Journal of
Experimental Psychology

VOL. 25, NO. 6 DECEMBER, 1939

TIME ORDER ERROR IN SUCCESSIVE JUDGMENTS AND IN REFLEXES. I. INHIBITION OF THE JUDGMENT AND THE REFLEX [1]

BY HELEN PEAK

Randolph-Macon Woman's College

There is great diversity of opinion regarding the most fruitful sources of scientific observations. Some consider that thorough explorations must be completed within narrowly delimited fields of behavior before it is profitable to cut across the conventional and often arbitrary boundaries which stereotypes of thought and language have reared between the facts relating to different reaction situations. Though this more circumscribed approach has its advantages and should not be neglected, there would seem to be justification in supplementary attempts to discover general principles in widely different kinds of behavior.

The present study has undertaken to find some such general principles in lid responses to paired tone stimuli and in the absolute loudness judgments of the same stimuli made by the same subjects, and in turn to relate these principles to other reflex functions reported in the literature of physiology. This program is believed to derive its justification from several considerations. From a systematic point of view, any suc-

[1] The author wishes to thank the Institute of Human Relations, Yale University, for the loan of part of the apparatus which made this study possible. Certain of the results were reported at the meeting of the American Psychological Association, September, 1937, under the title 'Some Comparisons between Reflex Inhibition and Summation and the Time Error Function.'

cessful attempt to organize a large number of different phenomena under a small number of generalizations tends toward simplification of knowledge, which is the job of science. It should, then, be of considerable value to know to what extent any sorts of behavior, such as judgments and reflexes, are similarly related to specific experimental variables.

The choice of the reflex and the judgment for study has been made on more specific grounds than this, however. The choice has been determined first of all by the fact that it has been possible to set up a convenient situation in which both the lid reflex and the loudness judgment may be elicited by the same stimuli and plotted as functions of the same independent variables. In the second place, the facts about both types of activity, though not as yet brought into relation, provide suggestive evidence of similar functions within these fields. It has long been recognized, for example, that when a series of judgments is made upon the intensity of one stimulus of a pair relative to the other member of the pair, the judgments are not always equally divided between the categories of 'greater' and 'less,' even though the stimuli themselves are distributed symmetrically around a standard. This phenomenon, the familiar time order error, is said to be *positive* if 'less' judgments of the second stimulus of the pair predominate, and *negative* when the 'greater' are more frequent. Though it has been generally assumed that the 'error' depends in part upon successive presentation of stimuli, attention has been called of late to experiments in which unsymmetrical distributions of judgment result when the method of single stimuli is employed (**18**), a result which is said to suggest that comparison of *paired* stimuli is not essential to the time order error. This evidence rules out the possibility of regarding time order as the only determining factor, but neither on logical nor empirical grounds does it remove the probability that where stimuli are arranged in pairs, the preceding stimulation helps determine the characteristics of the response following. The time order problem continues to have important implications for judgments of paired stimuli.

This relationship of judgment to temporal order of stimuli

finds numerous parallels in the literature of physiology and physiological psychology, for in many different experiments upon responses ranging from electrical potentials of excised nerves to reflexes and even voluntary responses of the intact organism, it has been shown that response to a stimulus may be modified greatly by preceding stimulus-response events and that the nature of the modification depends in part on the interval elapsing between the stimuli. These parallels justify an expectation that comparisons between the two fields may prove fruitful.

Finally, it is of interest to compare reflex functions with higher level functions because of the somewhat doubtful relation of the data of reflex physiology and psychology to the field of psychology as a whole. It is, of course, possible to justify any functional fact in and for itself, but if the value of the knowledge of reflex functions does not extend beyond the mere statements of the facts themselves, then they do have little general psychological interest. If, however, variations in any kind of behavior in an organism can be shown to be accompanied by concomitant variation in another activity, the facts relative to each type of reaction take on added significance because of the possibility of prediction where events have been shown to be correlated. Furthermore, where variations in both activities show the same functional relation to certain varying determinants (independent variables) then, again, both sets of facts are enhanced in value, because the generality of the function in question is increased, as well as the generality of the hypothetical factors (neural activities, sets, etc.) which are assumed to intervene between the observed independent and dependent variables. So if reflexes and judgments do show similar functions, it appears likely that reflex studies may throw light on factors influencing other activities in the organism and vice versa. In this way they may become the source of important discoveries which might otherwise be overlooked, especially in the more complex activities. It is, therefore, of some interest to explore the similarities between lid reflexes and loudness judgments, remembering of course that if similar functions are discovered,

neither activity is to be thought of as in any sense an 'explanation' of the other, but simply as a special case of a general principle operating in the organism. On the other hand, in so far as the functions prove to be different, such facts will also serve to set in relief the critical and important differences in the different levels of behavior organization and in the intervening factors that cause these differences. Thus any activity in the organism may conceivably illuminate any other, as well as the functioning of the organism as a whole, in so far as their mutual similarities and differences are made definite by experimental test.

The set-up of the present experiment was such that the subjects judged paired tones in absolute terms.[2] The first stimulus of a pair varied in intensity from trial to trial. The second remained constant.[3] Single stimuli of the same intensities as the first of the paired stimuli were interspersed in the series and judged under the same conditions. The onset of the first stimulus was separated from the second by intervals of 177 or 285 ms. and each tone lasted 50 ms. Lid responses to these same tones were recorded photographically. The experimental results will be presented in two papers. Since the course of judgments of the *first* as well as the *second* stimulus of a pair should throw light on the origin of the time error, the two sets of data will be analyzed separately and then studied from the point of view of their relationship. The present paper reports a time order error in both judgments and reflexes, and then considers the judgments and reflexes of the second tone of the pair as functions of various variables. A second paper will undertake to discuss the reactions to the first stimulus of a pair, the interrelations between these reactions and between R's and J's. Finally, in a third article

[2] This can hardly be called the method of single stimuli, nor with complete justification, the method of absolute judgment, for the stimuli were paired and there is good reason to believe that comparison is made between them. As a consequence, the judgments are relative, though they are stated in absolute terms. For a discussion of terminology see Volkmann (27).

[3] Hereafter S_1 will be used to refer to the first stimulus of a pair, S_2 to the second stimulus, R_1 to the reflex to S_1, R_2 to the reflex to S_2, J_1 to the judgment of S_1, J_2 to the judgment of S_2.

various time error theories will be considered in their relation to the present data.

In the present paper, data are presented on the following points: (1) *The presence of a positive time order error.* It is interesting to note the direction of the time error at the short intervals used in this experiment, inasmuch as Köhler (**13**) has suggested that there is no time error or a very slightly negative one at short intervals ($\frac{1}{2}$ and 1 sec.). His results were, however, admittedly ambiguous. Furthermore, Müller (**16**) long ago suggested that the time error appeared only where the differences between stimuli were small. The usual methods have not permitted a check of this statement because if, as is commonly the case, the judgments are given in relative terms, the effect of the first stimulus on the judgment of the second is prevented from appearing as soon as the difference between the stimuli becomes super-threshold. The present method makes possible the observation of the existence of such effects between stimuli which are widely separated on the intensive scale.

(2) *The Judgments of, and Reflexes to, the Second Stimulus as Functions of the Intensity of the First Stimulus.*—This is an independent variable which has been shown to be related to inhibition and reinforcement processes in spinal reflexes (**25**) and in reflexes of the intact organism (**19**). In the time error literature the question of the effect of intensity seems never to have been asked in just this way. The chief interest in stimulus intensity as an independent variable has centered in the problem of change in the direction of the time order error produced by following a *series* of judgments of stimuli at an intense level by a *series* of judgments of weak stimuli (**23**). Still other writers have been concerned with the effect of the interpolation of backgrounds of various intensive levels between the first and second stimuli (**24, 15**). The setting of Heymans' law (**7**) is more closely related to the present experiment, although it was stated with reference to a change in sensory thresholds of one stimulus as a function of changing intensity of another stimulus, while the dependent variable of the present study is a judgment in absolute terms of paired

stimuli separated by a supra-liminal difference. The independent variable, stimulus intensity of a preceding or concomitant stimulus, is the same, however. The literature on masking and auditory fatigue may also include comparable situations (**26**, Ch. 8), as do certain studies which have sought to show a relationship between the differential threshold and the amount, intensity, and exposure time of the initial stimulus of a pair.[4]

(3) *Judgments of S_2 and Reflexes to S_2 as a Function of the Interval between S_1 and S_2.*—As suggested above, this variable has been of interest in many behavior situations. In time error work it has been called the p function (**17**) and has been studied by a number of investigators who have found the time error changing with increase in the interval between stimuli.[5] The relation of time interval to recovery from refractory phase in nerve (**6**), to the facilitation and inhibition in reflexes of physiological preparations (**3**), and to reflexes of the intact organism is well known (**8**). In fact, it has become increasingly apparent that the effect of a single independent variable such as stimulus intensity, cannot be considered apart from dynamic changes of inhibition, excitation, and recovery which follow upon stimulation.

Apparatus

Tones of 1000 cycles were delivered to the subject through headphones of the balanced armature type (Baldwin, Type C). Sponge rubber pads between ears and phone prevented discomfort and confined the tone to the ear canal. An alternating current was produced by means of a 1000 cycle tuning fork oscillator (General Radio, Type 213B) actuated by a 6 volt battery. The oscillator was enclosed in a soundproof box and ran throughout the experiment. Rheostats in the circuit made it possible to keep constant the input to the oscillator. Figure 1 represents the apparatus used for controlling the intensity of the two tones. K_1 and K_2 are the contacts of a double pole, single throw, make-on-break relay (Western Electric, E-111) which controlled the circuit through potentiometer X.[6] S_1 and S_2 are contacts of a similar relay controlling circuit Y. The circuits through X and Y were never closed at the same time. The relays were operated by means of pendulum switches. The current through relay magnets (not pictured here) was kept constant throughout the experiment.

[4] See, for example, Holway, Golding, and Zigler (**9**). This is only one of a series of these papers.

[5] See, for example, Needham's summary (**18**).

[6] Double breaks were used in the circuits to avoid tone from induced current in the phones.

By means of this arrangement, variation in the setting of potentiometer *Y* produced variations in the intensity of the first tone of the pair, S_1. The setting of potentiometer *X*, controlling the second tone, was kept constant at the zero setting; *i.e.*, with no potentiometer resistance in the headphone circuit. Four different intensity values were used for the first stimulus (S_1). The settings for these intensities were determined in the following manner. With the potentiometer at zero and circuit closed, the input current was adjusted so that 6 volts were delivered through an output meter placed across the headphones. The amount of resistance was then increased to a point at which the tone was always audible.[7] This point was marked 'intensity 1' on the dial. Since at this point the voltage was too low to be read from the voltmeter (G. E. Outputmeter 483), a 100 volt-60 cycle source was substituted for the 6 volt source for calibration purposes. Under these conditions the intensity 1 position showed

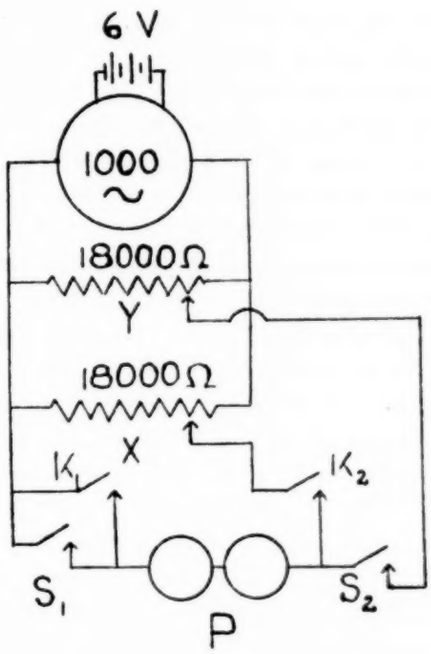


FIG. 1. Oscillator circuit.

a reading of 0.3 volts. By moving the dial, this value was increased sevenfold to the second intensity position.[8] This second value was in turn multiplied by seven to produce the third intensity, and similarly the fourth position was approximately (slightly less than) seven times the voltage of the third. The potentiometer dial was marked at the four positions producing these readings. The voltage values of these settings are given in Table 1. The values increase geometrically but it is not certain that the tones which reached the subjects bore precisely this relation to each other, because of possible changes in inductance effect within the earphones over the wide

[7] This is not to be taken as an accurate threshold value, for the potentiometer was such that the intensity of the stimulus could not be adjusted in small enough steps to make possible the accurate determination of the absolute threshold.

[8] The multiple seven was chosen simply as a convenient number enabling us to include the range of the potentiometer in four intensity steps. It was necessary to include large intensities because of the high threshold for the reflex.

TABLE 1

Stimulus	Meter-reading with 100 V. Source	Voltage with 6 V. Source	Voltage Squared
1	0.3	0.018	0.000324
2	2.1	0.126	0.015876
3	14.7	0.882	0.777924
4	100.0	6.00	36.000000

intensity range used here. These unmeasured changes in inductance would vitiate the assumption of a linear relation between voltage and amperage and consequently introduce a small error into the assumption that voltage is proportional to energy delivered by the phones (12).

The duration of the stimuli and the intervals between them were controlled by the use of a pendulum which struck successively four rocking contacts of a modified Dodge pendulum photochronograph (20). Opening the first contact broke the circuit through the relay magnets and closed the oscillator circuit through potentiometer Y; the second pendulum contact closed the relay circuit and broke the oscillator circuit Y; the third closed the oscillator circuit through potentiometer X, and the fourth broke the same circuit. This made it certain that circuits X and Y were never closed at the same time. The calibration of the duration of the stimuli and of the intervals between them was made by means of an optical lever system, fastened directly to the headphone diaphragm for photographic recording. These calibrations were made at the beginning of the experiment, at the end, and at a point about half way through, between the second and third subjects. The average duration of S_1 was 51.2 $\pm$ 3.2 ms.; of S_2 (at short interval), 52.57 $\pm$ 2.9 ms.; of S_2 (at long interval), 52.42 $\pm$ 2.7 ms. The differences between these durations are in no cases reliable. The intervals between onset of S_1 and S_2 averaged 177.3 $\pm$ 2.5 ms. and 285.0 $\pm$ 1.8 ms. These same photographic records provided the opportunity of studying the characteristics of the onset of the stimulus. There was no audible click and, for the most part, the records showed a smooth onset with about 5 ms. required for the tone to reach maximum. There was a slight initial fluctuation in amplitude lasting about 5 ms. after the maximum amplitude was reached. This may have been due to a slight bouncing of the relay contacts which, though not great enough in extent to break the circuit and cause a click, may have produced a small variation in the resistance through the contacts. It may also have resulted from an overthrow of the recording mirror.

The lid responses were recorded by the shadowed lash method (5) on a photographic paper carried by the same pendulum which controlled stimulus intervals and durations. Amplitudes of response were read to the nearest half mm. Responses were magnified between 3 and 4 times and the data are reported in terms of the magnified R's.

PROCEDURE

As pointed out above, the intensity of S_2 remained the same throughout the experiment and was equal to 4, the largest value of S_1. The four intensities of S_1 were presented in a carefully counterbalanced random order, each stimulus intensity appearing in each 5 trials and each having approximately equal average positions in the daily series. A single unpaired stimulus was also included within each 5 trials, the intensity and position varying in random order. S_2 (always intensity 4) was also in-

cluded as one of the single stimuli. The trials were arranged in banks of 25, including 20 paired and 5 unpaired stimuli. The intervals between stimuli remained the same within a single bank of 25. It was changed at the end of each 25 trials and each day began with a different time interval. The average position of these banks is equal if the results of all subjects are averaged. The subjects reported no insight into the order of the stimulus intensities. Three of the subjects came for 5 sittings. Two, D and S, served for 8 and 9 sittings respectively.

None of the subjects knew anything about the purpose of the experiment. After having been adjusted in the apparatus, each subject, except D, was given the following instructions: "When you hear the shutter open, this will be your signal, (1) to fixate the cross in front of you; (2) to prepare to judge the loudness of the tones which will follow. As soon as you have heard the tones, state the loudness of each as one of the following values: 0, 1, 2, 3, 4, 5, 6, 7, 8, or 9. Use 9 for the loudest tone which you hear in the headphones, 1 for the weakest, and 0 for no tone. Reserve number 5 for that tone intensity falling directly in the middle of the series. The remaining numbers should represent intensities falling in between this and the extremes. You will always be able to hear the second tone; occasionally you may not be able to hear the first. Always give your report on the first tone first, and on the second tone second. The differences in the second tone are so small that you will have to listen carefully." Subject D's instructions suggested a 7 step scale instead of the 9 steps and in the first half of the experiment she judged only the second tone of the pair. Also, in the first half of her experiment, no single tones were used. In other respects, the procedure was the same as that for other subjects. Five women college students served as subjects.

After the instructions were understood, a series of 20 single tones was given to the subject to indicate the intensity range within which judgments were to be made. The subject judged these tones which included 5 instances of each of the 4 intensities used in the main experiment, presented in random order. The experiment proper then began. The shutter of the apparatus was opened 2–3 sec. before the pendulum was noiselessly released, setting off S_1 and S_2 of that trial. Judgments were given and set down at once on the back of the photographic record of the lid responses to the pair of tones. The next trial followed 45–60 sec. after the beginning of the first trial.

Results

1. *Positive Time Order Errors*

Table 2 contains the individual averages (M), sigmas (σ), and number of cases (N), of the loudness judgments of S_1 and S_2,[9] where both have intensity value 4. It is apparent from this Table that for 3 of the subjects the second stimulus tended to be judged smaller than the first when equal to it in intensity, particularly at the short interval. This is equivalent to saying that a stimulus of a given intensity, frequency, and duration is judged greater when first in a pair

[9] Medians were calculated for each subject but are not reported as the results are closely like averages without a single reversal in direction.

TABLE 2

LOUDNESS JUDGMENTS OF S_1 AND S_2 WHEN $S_1 = 4$

Subject			B	W	S	E	D_2*
Interval		*M*	8.9	8.0	7.5	7.1	3.5
	J_1	*N*	44	49	75	41	31
177		σ	0.3	1.1	1.2	0.9	1.8
ms.		*M*	7.4	8.0	5.4	8.1	2.7
	J_2	*N*	44	49	75	41	31
		σ	0.8	1.2	1.5	0.3	0.7
		M	9.0	8.5	8.4	8.1	4.0
285	J_1	*N*	47	46	73	35	32
		σ	0.3	0.9	0.6	0.8	0.8
ms.		*M*	8.1	8.4	7.7	8.4	3.9
	J_2	*N*	47	46	73	35	32
		σ	0.5	1.2	1.1	0.7	0.6

* D was judging on a seven step scale while others used nine steps. D_1 is omitted because judgments of only S_2 were made (see text, p. 545). D_1 refers to data from first half of experiment on subject *D*.

than when second, which is by definition a positive time order error.

When these average differences between J_1 and J_2 (with S_1 equal to S_2) are examined for consistency it is found that at 177 ms. for subject B, $J_1 > J_2$ 100 percent of the total trials; for subject S, 89 percent of the trials; for D, 97 percent; for W, 11 percent (with 71 percent equal); for E, 23 percent (with 10 percent equal). At 285 ms., $J_1 > J_2$ for B, 77 percent of the trials; for S, 46 percent (with 41 percent equal); D, 42 percent (with 42 percent equal); for W, 13 percent (with 82 percent equal); for E, 0 percent (with 44 percent equal). Thus there is a preponderance of 'greater' over 'less' for B, S, and D at both intervals and for W at 285 ms.

Treated in the conventional manner of time order errors, these results appear in Table 3.[10] It will be seen that E at both intervals and W at 177 ms. are the only instances where a positive time error is not present. For the other subjects the positive time error is marked and tends to be more

[10] The data of Table 3 are calculated by the formula:

$$\frac{\#\text{ times } J_1 > J_2 - \#\text{ times } J_2 > J_1}{N}.$$

positive at the shorter interval. This is in contrast with Köhler's failure (**13**) to obtain positive time errors below $1\frac{1}{2}$ sec. intervals and with Müller's statement (**16**) that they occur only where differences between stimuli are small. Furthermore, the effect of position is not limited to the situation where S_1 and S_2 are equal, although the effect cannot be stated in the usual manner when S_1 is consistently judged smaller than S_2. The absolute values show that following all intensities of S_1, S_2 is judged smaller than the same stimulus in the first position. This is true of all subjects except D at 177 ms. at the smallest intensity of S_1[11] and subject E for all values at the 177 ms. interval. W shows the effect at both intervals and E

TABLE 3

TIME ERRORS WHEN $S_1 = S_2$ = INTENSITY 4

Subject	177 ms.	285 ms.
B	+100	+59
W	− 8	+ 8.8
S	+ 88	+33
E	− 42	−48
D_2	+ 93	+26

at 285 ms., even though they did not show a positive time error when S_1 and S_2 were equal.

2. *Inhibition as a Function of Intensity* S_1

a. Judgment of S_2.—Table 4 contains the average judgments of the second stimulus following the four intensities of S_1 at 177 and 285 ms. intervals. In addition to the fact of decrement in J_2 noted above, it is evident that this decrement bears a relationship to the intensity of S_1 which is consistent for all subjects except E. In general the size of J_2 decreases as the intensity of S_1 increases until S_1 reaches the largest intensity. Though inhibition is present at this point, the function is reversed toward increase for all subjects except *D* at 177 ms. This will be seen in the averages of Table 4 and in Fig. 2.

[11] This subject failed to hear S_1 and S_2 as separate at this intensity. Failure to obtain inhibition of J_2 here is undoubtedly related to this fact.

TABLE 4

AVERAGE VALUE LOUDNESS JUDGMENTS OF S_2

Subject		J_2—177 ms.				J_2—285 ms.			
	Intensity S_1 →	1	2	3	4	1	2	3	4
B	*M*	7.7	7.2	6.4	7.4	7.5	7.1	6.4	8.1
	N	47	47	48	44	47	47	49	47
	σ	0.2	0.8	1.4	0.8	0.7	0.7	0.8	0.5
W	*M*	7.1	6.6	6.0	8.0	7.2	7.0	6.4	8.4
	N	48	46	48	49	49	48	49	46
	σ	1.9	1.8	1.8	1.2	1.7	1.7	1.8	1.2
S	*M*	7.8	6.6	4.8	5.4	7.4	6.6	5.6	7.7
	N	73	72	75	75	72	73	74	73
	σ	1.0	1.1	1.3	1.5	1.1	1.2	1.6	1.1
E	*M*	7.1	7.1	7.4	8.1	6.0	5.8	6.6	8.4
	N	40	40	40	41	34	35	37	35
	σ	2.3	2.1	1.4	1.3	0.9	1.5	1.6	0.7
D_1	*M*	5.7	4.8	4.1	4.0	5.4	4.4	4.1	5.1
	N	32	31	33	31	24	24	23	24
	σ	0.6	0.9	0.6	0.6	0.7	0.6	0.7	0.8
D_2	*M*	4.3	3.6	2.8	2.7	3.9	3.6	2.9	3.9
	N	31	30	30	31	32	31	31	32
	σ	0.6	0.6	0.6	0.7	0.7	0.5	0.5	0.6

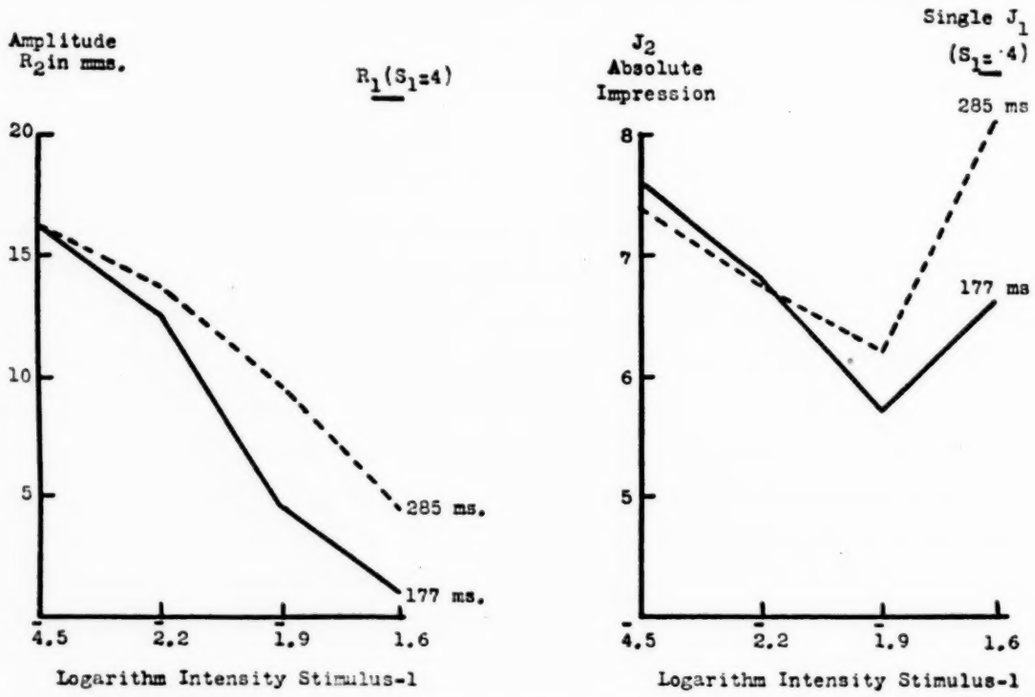


FIG. 2. Reflexes and judgments as functions of stimulus intensity. Average values of R_2's, all subjects included (subject S only 8 cases, see text). Average values J_2 with D omitted because judgments were on a different scale.

The critical ratios of the difference between averages have been calculated but are not reported in detail because the distributions are not normal, which leaves in doubt the interpretation of the values.[12] Since the differences cannot be evaluated in the usual way, Table 5 is presented to indicate

TABLE 5

COMPARISON SESSION AVERAGES OF J_2's FOLLOWING DIFFERENT INTENSITIES OF S_1

	Intensities of S_1 preceding J_2 ↓	Subjects					
		B	W	S	E	D_1	D_2
	*$J_2(1) > J_2(2)$	90%	50%	100%	50%	100%	100%
	$J_2(1) = J_2(2)$	0%	20%	0%	25%	0%	0%
	N	10	10	16	8	8	8
177 ms.	$J_2(2) > J_2(3)$	90%	80%	100%	37.5%	87.5%	100%
	$J_2(2) = J_2(3)$	0%	10%	0%	12.5%	0%	0%
	N	10	10	10	8	8	8
	$J_2(3) < J_2(4)$	90%	100%	81.3%	100%	25%†	37.5%†
	$J_2(3) = J_2(4)$	10%	0%	0%	0%	25%	12.5%
	N	10	10	16	8	8	8
	$J_2(1) > J_2(2)$	70%	70%	87.5%	43%	50%	75%
	$J_2(1) = J_2(2)$	0%	0%	0%	28.6%	25%	12.5%
	N	10	10	16	8	8	8
285 ms.	$J_2(2) > J_2(3)$	80%	70%	100%	15%	87.5%	87.5%
	$J_2(2) = J_2(3)$	0%	10%	0%	14%	12.5%	12.5%
	N	10	0	16	8	8	8
	$J_2(3) < J_2(4)$	100%	90%	93.8%	100%	100%	100%
	$J_2(3) = J_2(4)$	0%	0%	6.2%	0%	0%	0%
	N	10	10	16	8	8	8

* To be read: J_2 following S_1 of intensity 1 is greater than J_2 following S_1 of intensity 2 90 percent time for subject *B*.

† The averages of Table 4 also indicate that J_2's following S_1 of value 3 are greater than those following S_1 of value 4; *i.e.*, this is consistent with that Table.

that even when the data are fractionated the tendencies are the same as those indicated by averages. Each session average of J_2 following each intensity of S_1 (consisting of 4 or 5

[12] The average of the ratio, $\frac{D}{\sigma d}$, for the 4 subjects showing results in the same direction (B, W, D, S) are: at interval 177 ms., difference between S_1 (1 and 2), 4.1; S_1 (2 and 3), 4.4; S_1 (3 and 4), 3.5; at 285 ms., S_1 (1 and 2), 2.6; S_1 (2 and 3), 3.8; S_1 (3 and 4), 6.6. The values are reliable, *i.e.*, 3 or greater, for all individual subjects except *W*, at all intervals and intensities.

trials) has been calculated and comparisons made between these sub-averages of J_2. The Table contains the percent of the sub-averages of J_2 following a given intensity of S_1 which are greater or less than J_2 sub-averages following an adjacent intensity. The percent of times the values are equal, as well as the number of sub-averages (N), is also given. In this Table, row 1 is to be read: For subject B, 90 percent of the experimental sessions showed average judgments of the second stimulus larger following a first stimulus of intensity 1 than following a first stimulus of intensity 2. For subject W, 50 percent of J_2 sub-averages following a first stimulus of intensity 1 were greater than those following a first stimulus of intensity 2, etc. In the second row, for subject B, 0 percent J_2 sub-averages following S_1 of value 1 were equal to those following an S_1 of value 2, etc. Despite this extreme fractionation, the direction of the differences indicated by a comparison of the averages of Table 4 are highly consistent with the comparisons of the sub-averages of Table 5. Only subject W in one comparison (of J_2's following S_1's of 1 and 2 at 177 ms.), subject D in one comparison (J_2's following S_1's of 1 and 2 at 285 ms.), and subject E, whose averages were wholly inconsistent by all criteria, showed 50 percent or more of the comparisons of sub-averages taking a different direction from that indicated by comparison of averages of total data in Table 4. This suggests that the averages of Table 4 give a fair picture of the actual tendencies indicated by individual trials.

b. Reflexes to S_2.—The data on the reflexes are found in Table 6. Here are stated the average amplitudes, sigmas and number of cases of the reflex to the second stimulus as a function of the intensity of the first stimulus. Table 6 and Fig. 2 indicate that R_2, like J_2, decreases in size as the intensity of S_1 increases. There is, however, no evidence of a terminal reversal in the curve, such as is seen for J_2. The tendency is consistent for every subject, though the differences between R_2's following S_1 at intensity 1 and at intensity 2 are not in every case statistically reliable. The differences are greater between the other values.

Again, the distributions are not normal and critical ratios not interpretable in the usual manner.[13] Table 7 reveals the consistency of the R_2's by comparing R_2 sub-averages following each intensity of S_1 and giving the percent of these comparisons in which the R_2's under one condition exceed those

TABLE 6

AMPLITUDE OF REFLEXES TO S_2 (R_2's)

Subject		177 ms.				285 ms.			
	Intensity S_1→	1	2	3	4	1	2	3	4
B	*M*	12.7	12.6	3.9	2.2	12.6	11.5	9.7	4.1
	N	40	43	35	22	45	41	43	36
	σ	10.9	10.8	9.8	3.5	7.0	7.3	6.1	3.5
W	*M*	16.0	14.3	5.5	1.7	17.4	14.9	11.0	7.6
	N	48	47	48	44	48	47	47	47
	σ	12.0	11.3	2.6	1.1	13.1	11.1	7.5	3.6
S*	*M*	39.4	22.7	5.9	0.1	29.8	31.2	16.0	1.8
	N	8	7	8	7	7	5	6	4
	σ	19.5	7.6	4.5	0.2	12.8	10.0	7.6	1.4
E	*M*	26.6	21.0	8.1	0.3	19.1	12.6	9.3	3.9
	N	33	33	37	40	29	24	30	39
	σ	16.6	16.1	6.9	0.1	22.7	14.9	8.8	5.1
D_1	*M*	11.2	6.8	1.8	0.3	9.8	7.9	4.0	1.0
	N	32	31	33	31	31	32	31	29
	σ	5.6	4.6	2.1	1.7	6.1	4.2	2.2	0.6
D_2	*M*	17.0	11.3	4.4	0.8	19.0	16.9	10.9	2.6
	N	31	30	29	24	32	31	31	31
	σ	8.1	5.1	3.8	3.6	10.1	7.5	7.2	4.0

* The *N* for this subject is very small because by the end of the first day she showed a tendency to develop anticipatory responses which interfered with accurate readings of R_2.

under another. *N* takes the same values as in Table 5.

PAGES WAVY

trials) has been calculated and comparisons made between these sub-averages of J_2. The Table contains the percent of the sub-averages of J_2 following a given intensity of S_1 which are greater or less than J_2 sub-averages following an adjacent intensity. The percent of times the values are equal, as well as the number of sub-averages (N), is also given. In this Table, row 1 is to be read: For subject B, 90 percent of the experimental sessions showed average judgments of the second stimulus larger following a first stimulus of intensity 1 than following a first stimulus of intensity 2. For subject W, 50 percent of J_2 sub-averages following a first stimulus of intensity 1 were greater than those following a first stimulus of intensity 2, etc. In the second row, for subject B, 0 percent J_2 sub-averages following S_1 of value 1 were equal to those following an S_1 of value 2, etc. Despite this extreme fractionation, the direction of the differences indicated by a comparison of the averages of Table 4 are highly consistent with the comparisons of the sub-averages of Table 5. Only subject W in one comparison (of J_2's following S_1's of 1 and 2 at 177 ms.), subject D in one comparison (J_2's following S_1's of 1 and 2 at 285 ms.), and subject E, whose averages were wholly inconsistent by all criteria, showed 50 percent or more of the comparisons of sub-averages taking a different direction from that indicated by comparison of averages of total data in Table 4. This suggests that the averages of Table 4 give a fair picture of the actual tendencies indicated by individual trials.

b. Reflexes to S_2.—The data on the reflexes are found in Table 6. Here are stated the average amplitudes, sigmas and number of cases of the reflex to the second stimulus as a function of the intensity of the first stimulus. Table 6 and Fig. 2 indicate that R_2, like J_2, decreases in size as the intensity of S_1 increases. There is, however, no evidence of a terminal reversal in the curve, such as is seen for J_2. The tendency is consistent for every subject, though the differences between R_2's following S_1 at intensity 1 and at intensity 2 are not in every case statistically reliable. The differences are greater between the other values.

Again, the distributions are not normal and critical ratios not interpretable in the usual manner.[13] Table 7 reveals the consistency of the R_2's by comparing R_2 sub-averages following each intensity of S_1 and giving the percent of these comparisons in which the R_2's under one condition exceed those

TABLE 6

AMPLITUDE OF REFLEXES TO S_2 (R_2's)

Subject		177 ms.				285 ms.			
	Intensity S_1→	1	2	3	4	1	2	3	4
B	*M*	12.7	12.6	3.9	2.2	12.6	11.5	9.7	4.1
	N	40	43	35	22	45	41	43	36
	σ	10.9	10.8	9.8	3.5	7.0	7.3	6.1	3.5
W	*M*	16.0	14.3	5.5	1.7	17.4	14.9	11.0	7.6
	N	48	47	48	44	48	47	47	47
	σ	12.0	11.3	2.6	1.1	13.1	11.1	7.5	3.6
S*	*M*	39.4	22.7	5.9	0.1	29.8	31.2	16.0	1.8
	N	8	7	8	7	7	5	6	4
	σ	19.5	7.6	4.5	0.2	12.8	10.0	7.6	1.4
E	*M*	26.6	21.0	8.1	0.3	19.1	12.6	9.3	3.9
	N	33	33	37	40	29	24	30	39
	σ	16.6	16.1	6.9	0.1	22.7	14.9	8.8	5.1
D_1	*M*	11.2	6.8	1.8	0.3	9.8	7.9	4.0	1.0
	N	32	31	33	31	31	32	31	29
	σ	5.6	4.6	2.1	1.7	6.1	4.2	2.2	0.6
D_2	*M*	17.0	11.3	4.4	0.8	19.0	16.9	10.9	2.6
	N	31	30	29	24	32	31	31	31
	σ	8.1	5.1	3.8	3.6	10.1	7.5	7.2	4.0

* The *N* for this subject is very small because by the end of the first day she showed a tendency to develop anticipatory responses which interfered with accurate readings of R_2.

under another. *N* takes the same values as in Table 5. There were no equal average R_2 values. It is clear that the tendency indicated by the general average is highly consistent with the sub-averages in most cases and in all subjects the percent of comparisons of sub-averages taking the same

[13] The average ratio, $\frac{D}{\sigma d}$, for the five subjects are: at 177 ms., S_1 (1 and 2), 1.9; S_1 (2 and 3), 4.3; S_1 (3 and 4), 4.7; 285 ms., S_1 (1 and 2), 1.1; S_1 (2 and 3), 1.9; S_1 (3 and 4), 2.6.

TABLE 7

COMPARISON SESSION AVERAGES OF R_2 FOLLOWING DIFFERENT INTENSITIES OF S_1

	Intensities of S_1 preceding R_2	Subjects *				
		B	W	E	D_1	D_2
177 ms.	† $R_2(1) > R_2(2)$	60%	80%	100%	100%	100%
	$R_2(2) > R_2(3)$	100	90	100	100	100
	$R_2(3) > R_2(4)$	90	100	100	87	87
285 ms.	$R_2(1) > R_2(2)$	90	70	80	100	64
	$R_2(2) > R_2(3)$	60	90	100	100	100
	$R_2(3) > R_2(4)$	80	90	100	100	100

* Values for subject S are not given because of early appearance of anticipatory R's distorting size of R_2 and reducing N.

† To be read: R_2 following S_1 of intensity 1 is greater than R_2 following S_1 of intensity 2 60 percent time for subject *B*.

direction as comparisons of averages of Table 6 are greater than 50 percent (chance).

3. *Inhibition as a Function of the Interval Between S_1 and S_2*

As already pointed out, J_2 is, as a rule, inhibited at both intervals used in this experiment, with a tendency for the inhibition to decrease in amount at the longer interval and larger intensities of S_1. At the two lowest intensities used, however, the curves cross for 4 of the 5 subjects, indicating slightly more inhibition at the longer interval. Though this difference is small and disappears in the average (see Fig. 2), there is clearly less difference in the amount of inhibition at 177 and 285 ms. following small than following large intensities of S_1. This suggests a delay in the development of inhibition at low stimulus intensities, while at the larger intensities the second response shows marked signs of recovery. Subject W is the only one whose results do not show this trend. For her, all the values following low and high intensities of S_1 at 285 ms. exceed those at 177 ms. There is recovery from inhibition at the longer interval. In subject E there is more inhibition at 285 ms. than at 177 at all intensities except 4.

The comparison of the effect of time intervals on R_2 is incomplete because two subjects in particular, S and E, showed an early tendency for secondary lid responses to appear anticipatory to S_2. This resulted in such a reduction of the number of cases of R_2 which could be accurately measured and included in the average that the significance of the values is reduced. The results appear to be similar to those found for J_2, however. In all cases the average amplitude of R_2 at 285 ms. exceeds R_2 at 177 ms. when S_1 has intensity 3 or 4. At 1 and 2, the differences are unreliable but in some cases R_2 is actually smaller at the longer interval, suggesting again delay in the building up and disappearance of the inhibitory effect at the small intensities of S_1.

Discussion

A Generalized Statement of J_2 and R_2 Results

At this time no attempt will be made to discuss the applicability of the various theories of the time order error to these results. That will be undertaken when the data for the responses to the first stimulus have also been presented (**21**), for it is evident that the time error must be a function of judgments of both stimuli of a pair, whether these judgments are expressed in absolute or relative terms. The present task is to try to find some generalized description of the experimental data which will cover the facts presented. Such a description will be an important part of the account of time errors to be found in the third paper.

The method followed in the discussion of this section is one familiar to scientific theorizing. Certain relations between independent and dependent variables are established by experiment. Similar relations are observed to exist between the same independent variable and other dependent variables. These functions become the basis of an hypothesis that similar processes may be affecting both reactions and that other independent variables may also affect similarly the reactions in question. In the present comparisons, changes in amount of inhibition of J_2's, R_2's and certain reflexes of the spinal cord are seen to be related similarly to

stimulus intensity, but in spinal reflexes amount of inhibition is known to be complicated by factors such as rate of build-up, rate of dispersion, the point at which the second stimulus is introduced. It is a valid procedure, therefore, to suggest that these other aspects of inhibitory processes in J_2's and R_2's may also show relations to the independent variable similar to those discovered in spinal reflexes. Then it remains to be seen whether the experimental data on J_2 and R_2 fit in with such additional hypotheses. If they do, the conclusion may be drawn that the hypothesis in question is a tenable one, though other hypotheses are not necessarily disproved thereby.

Three assumptions regarding the direction of the relation between the intensity of the first stimulus of a pair and certain features of inhibitory effects which develop from that stimulus have been considered relevant to the results of this experiment. (1) Inhibition [14] increases with increase in the intensity of S_1. The question about nature of the curve representing this relationship will be discussed below. (2) Time to maximum inhibition decreases with increase in the intensity of S_1. (3) The time for dispersion of weak inhibitory effects is less than that for dispersion of strong effects. It is proposed to examine the extent to which these assumptions appear to be valid for J_2 and R_2 in the present experiment and for data from the physiological literature, and to discuss supplementary assumptions which become necessary in order to account for certain differences in R_2 and J_2.

The first assumption is the most obvious of the three, since the relation described is apparent in a portion of both curves of Fig. 2. The writer has already discussed elsewhere (19) the similarity between this function in the lid reflex, in spinal reflexes, and in perceptual data reported in the Heymans' law experiments. There reference was made to evidence that in spinal reflexes, inhibitory as well as excitatory effects are increased by increased intensity of stimulation

[14] It is unnecessary to postulate anything about the physiological nature of the inhibitory function and nothing is implied at this time beyond the observed facts of decrement in reflex amplitude and in judgment values.

(**25**) (3, Ch. VI). The discovery of a similar function in psychophysical judgments of the second of paired stimuli under the present conditions suggests a further extension of the same principle.[15] The question arises as to the exact nature of this relationship. Heymans (**7**) claimed a straight line relation between stimulus intensity and inhibition but this has not been borne out in subsequent research (**2**) (**11**). The empirical curves of Fig. 2 suggest that a logarithmic relationship is approximated, especially when allowance is made for possible effects of differing recruitment rates of inhibition, and in both Figs. 3 and 4 to be described below, inhibition has been assumed to increase directly as the logarithm of the stimulus intensity.

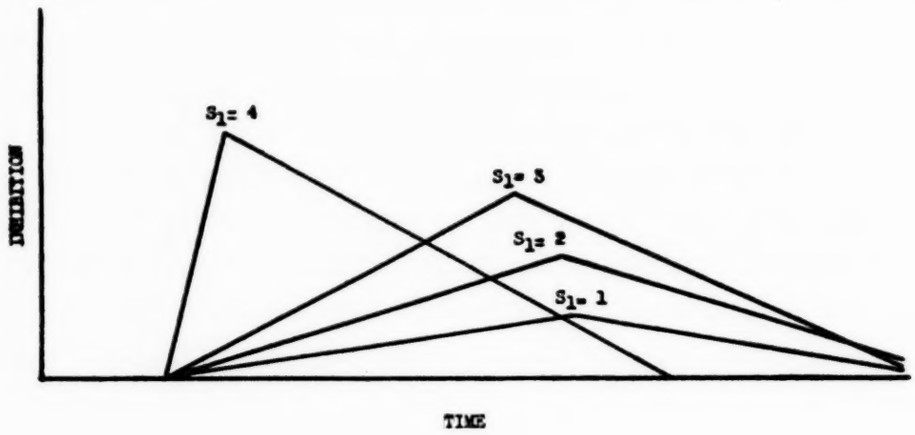
The second assumption, like the first, finds support in certain observations of reflex physiology. Sherrington and his colleagues (**3**, Fig. 50, pp. 98–99) have reported a decrease in the speed of attaining maximum relaxation of postural and of crossed-extensor reflexes with increase in the intensity of stimulation of ipselateral and contralateral fibres respectively.[16] A function of this sort is compatible in the present case with the appearance of increased inhibition at the longer interval between S_1 and S_2 where S_1 was weak, a thing which never occurred with the more intense stimuli.[17] There is no way of knowing the exact manner in which time to reach maximum inhibition is related to stimulus intensity, but one of the possible relationships is incorporated in Fig. 3 where time to maximum has been assumed to decrease directly as stimulus intensity increases. This assumption will be seen to be compatible with the terminal reversal in the J_2 curve, a fact which justifies its consideration. Another possible

[15] It is recognized that the complete range of intensities of S_1 and S_2 has not been studied and, obviously, the generalization must be limited to the conditions of the present experiment.

[16] Though the inhibition in the experiments cited resulted from stimulation of systems partially different from those producing the excitation, there is much evidence that summative inhibitory effects resulting from repetitive stimulation of the same afferent nerve are similar to those resulting from stimulation of another ipselateral afferent. (See, for example, **3**, p. 94.)

[17] Both latency and recruitment rate of inhibition may, of course, be involved in the 'time to maximum inhibition' and both may vary with stimulus intensity.

relationship of these variables will be discussed in connection with Fig. 4.



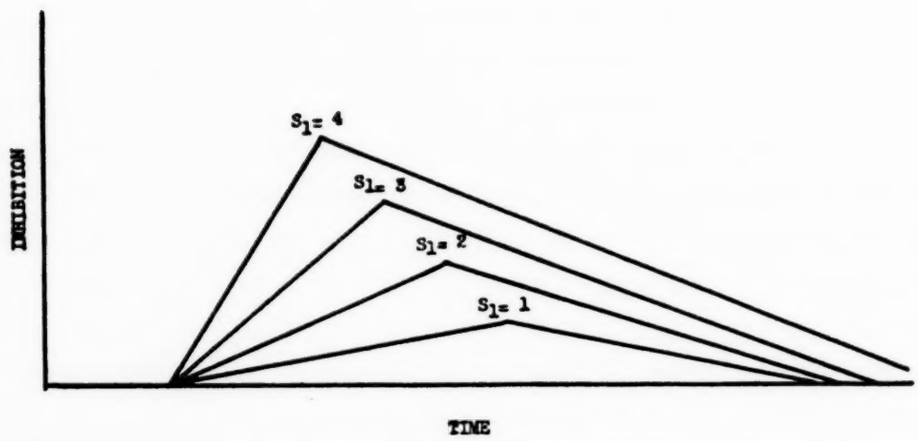


FIG. 3. Theoretical curves of the course of development of inhibitory processes. Amounts of inhibition are represented on the vertical axis, and time on the horizontal axis. The curve labelled, $S_1 = 4$, represents the course of developing inhibition processes following intensity 4 of the first stimulus of a pair; $S_1 = 3$, that following a stimulus of intensity 3, etc. This figure is based on the assumptions that inhibition increases as the logarithm of S_1 intensity, that time to maximum inhibition decreases directly as intensity of S_1, and that inhibitory effects disappear more slowly than they arise and that duration of dispersion of the effects increases directly as the stimulus intensity.

FIG. 4. Theoretical curves of the course of development of inhibitory processes. The axes represent the same variables as those of Fig. 3, but the curves are plotted with the assumption that time to maximum inhibition decreases as the logarithm of the stimulus intensity.

The present data afford little direct evidence for the third assumption, though the literature of physiology contributes

the relevant facts that inhibition following large stimuli endures longer than that following weak stimuli (**3**, p. 93) and that the rate of subsidence of inhibition is slower than that of recruitment (**3**, p. 87). These facts are considered relevant here because, where an inhibitory effect exists, the time value of its dispersion must be a factor in the results. Furthermore, it is necessary to take into consideration factors known to be related to any of the reactions under comparison and to see whether the operation of such factors would yield the results observed in the other reactions.

In Fig. 3, the three assumptions are brought together. Increasing amounts of inhibition are plotted on the ordinate and time on the abscissa.[18] The 4 curves of this figure represent the course of the rise and fall of inhibition set up by the 4 different intensities of the first stimulus as it should be reflected in changes in the responses to the second stimulus of a pair. The amount of inhibition is shown to increase as the logarithm of S_1, with the result that equal increments of inhibition are added from one intensity to the next, the stimuli being spaced geometrically. The time to maximum inhibition (*i.e.*, the distance from the ordinate to the peak of a curve) is plotted to decrease directly as the intensity of S_1 increases. As a consequence, the maxima of the curves are spaced at geometrically increasing intervals reading from right to left. Finally, the course of subsidence, the distance from the peak to the point at which a curve again crosses the abscissa, has been determined in the following manner. In accordance with the third assumption, the inhibition curve for S_1 of intensity 4 has been made to fall more slowly than it rises but fast enough to drop below the other curves in the right half of the figure. The other three curves of Fig. 3 also drop more slowly than they rise and the duration of the total time for dispersion in each case decreases directly as the

[18] No absolute time values have been assigned to the abscissa because it is certain that these values will vary somewhat from one neural level to another. See Lorento de Nó (**14**) for a discussion of differences in time values at cortical and reflex levels. Hilgard (**8**) has also drawn comparisons between spinal reflexes and reflexes in the intact organism.

stimulus intensity increases, so that dispersion of inhibition following weak stimuli takes less time than that following intense stimuli. Inhibition is shown to decrease at a steady rate in each curve. Though this is probably not a completely accurate representation, the particular configuration of the curves will not alter greatly their relative heights, which will be the object of consideration.

The question now arises as to how well the J_2 and R_2 results fit in with these assumptions. It has already been pointed out that at both time intervals between S_1 and S_2 the inhibition of J_2 increases with the intensity of S_1, except at intensity 4 where there is a reversal. This means that more inhibition resulted from intensity 3 than from intensity 2 and more from 2 than 1. These facts are wholly compatible with Fig. 3.

At intensity 4, however, the reversal is such that in some subjects a J_2 following this intensity was greater, *i.e.*, there was less inhibition, than in a J_2 following 3, while in other subjects, less inhibition of J_2 resulted from stimulus intensity 4 than 2. In subject W there was actually less decrement following the largest S_1 than after the very weakest. This result is consistent with Fig. 3 if the inhibition curve resulting from stimulus intensity 4 falls off earlier than the other curves as pictured, and if it be assumed that even though the curves of Fig. 3 represent relationships characteristic of all subjects, individuals may arrive at the different stages in development of inhibition in different absolute units of time. For example, at 177 ms., W would seem to have reached the stage in the course of inhibitory processes where the $S_1 = 4$ curve has dropped below the other curves, *i.e.*, where there is least inhibition following intensity 4. At this same interval, the process in subject D would seem not to have progressed beyond the point where curves $S_1 = 3$ and $S_1 = 4$ cross each other, so that for her alone the results show more inhibition at the large intensity than at any others. The other subjects may be similarly located in terms of the relative amounts of

inhibition which they show following intensity 4. That such processes should vary from individual to individual might well be expected in view of the well known individual differences in every sort of psychological and physiological reaction.

The reported effects on J_2 of the interval of time elapsing between S_1 and S_2 are also consistent with Fig. 3. All the subjects showed more inhibition at 177 than at 285 ms. following stimulus intensity 4. That is to say, by 177 ms., curve $S_1 = 4$ has apparently reached its crest and begun to drop for all subjects. At the weaker intensities, however, there was often more inhibition at 285 ms. This would be expected from Fig. 3 if the end of the 177 ms. interval were located to the right of the point where curve $S_1 = 4$ starts falling, as just suggested, but to the left of the point where the two weakest intensity curves begin to drop. Since the end of the 285 ms. interval must fall to the right of 177 ms., an increase in amount of inhibition at weak intensities would be expected in moving from short to longer time intervals until the point is reached where the inhibition curves have declined to the level attained at the shorter interval.

Turning to the R_2 data, it appears that the curves as plotted in Fig. 3 are adequate to cover the facts for the three smallest intensities, as in the case of J_2, for the relative values of R_2 following each intensity of S_1 are the same as those for J_2 at these intensities. Furthermore, similar effects of time interval on R_2 have been reported, with a tendency for inhibition to fall off at the long interval following the largest intensity of S_1 and a tendency for it to increase at this same interval following weak intensities. All these facts are covered by Fig. 3.

As pointed out, however, R_2 does not show the terminal reversal which characterizes the J_2 function, and for this reason it is obvious that although J_2 data following S_1 of value 4 fit the graph of Fig. 3, those of R_2 at this same value do not do so, for the amount of inhibition of R_2 does not drop below others at intensity 4 as it does for J_2. It is necessary, therefore, to discuss the possible reasons for this difference in

R_2 and J_2 functions.[19] The question arises, therefore, whether a new set of generalizations must be made for R_2. A point suggests itself which would make unnecessary any new assumptions. It will be shown in Table 6 of the second paper (**21**) that the threshold for reflexes is much higher than that for judgments, so that in a large proportion of trials there were no lid reactions to stimuli of the first and even of the second intensity. On the other hand, the tones were never below the perceptual threshold. It is certain, therefore, that although intensities of physical stimuli were identical, the maximum intensity of S_1 represented a great many more increments above threshold for judgments than for the reflex. If, then, the straightline inverse function of time to maximum inhibition and intensity holds for reflexes as well as for judgments, this would mean that an increase in intensity from 3 to 4 would be accompanied by a much greater absolute decrease in latency to maximum inhibition for the judgment than for the reflex, since these values are so much farther above the threshold of the reacting system. Consequently, in the reflex, the amount of inhibition resulting from an S_1 of value 4 would be less likely to fall below that from an S_1 of value 3 by the end of 177 or 285 ms. interval, than in the judgment. The curves for the reflex at highest intensity of S_1 would presumably have a relationship more like curves for $S_1 = 1$ or 2, the maximum points coming closer together in time than those for the judgment function. This would explain the difference in R_2 and J_2 curves at intensity 4 should the assumption prove valid.

[19] It should be mentioned that in an earlier paper (**19**) reporting a similar function of the lid reflex, a terminal rise in R_2 was noted at the large intensities of S_1, just as we have observed it here for J_2. In contrast to J_2, this tendency increased with practice and was definitely correlated with the frequency of responses anticipatory to S_2. In that experiment the fact that stimuli were presented to a given subject at one interval only appeared to facilitate the appearance of secondary responses in contrast to the present set-up where the conditioning process was relatively slow, perhaps because of alternating time intervals for the same subject. Had the present experiment continued, a reversal in R_2 function when S_1 was 4 would have undoubtedly occurred just as it did in the earlier experiment.

Alternative Generalizations Regarding R_2 and J_2

There are, of course, other possible relations between the variables of Fig. 3 as well as other possible explanations of the differences in R_2 and J_2 functions at the large intensity of S_1. It might be assumed, for example, that the time to maximum inhibition decreases as the logarithm of intensity S_1, instead of directly with the intensity, and a marked difference in the relative values of the curves of Fig. 3 would result from this change.[20] Figure 4 has been drawn exactly like Fig. 3 with this one assumption changed. As a consequence, the peaks of the inhibition curves of Fig. 4 are spaced equidistantly. It would have been possible, of course, to draw this figure so that the $S_1 = 4$ curve would have fallen below the $S_1 = 3$ curve as in Fig. 3. In this case, the figure would fit the facts of terminal reversal in J_2 at intensity 4 but all the other curves would have had to cross each other later on (*i.e.*, $S_1 = 3$ would fall below $S_1 = 2$ and $S_1 = 2$ below $S_1 = 1$). Since the data indicate that this does not occur within the limits of the present experiment and that neither the data of R_2 or J_2 would fit such a graph, it is nearer the facts to draw the inhibition curves of $S_1 = 1$, 2, and 3, as they are in Fig. 4, so that they do not cross within the interval studied. This means, then, that the curve of $S_1 = 4$ will not cross the others and that inhibition following this intensity will be greatest of all at all time intervals. In this case all the data which fitted Fig. 3 fit these curves except that now the observed J_2 values following intensity 4 are all too high for the curve. The R_2 data do fit, however, and it becomes necessary to discuss whether there are other possible explanations for the course taken by the J_2 values.

In the first place, it might be suggested that Fig. 4 is really a correct representation of J_2 and R_2 functions and that the apparent decrease in inhibition of J_2 following intensity 4

[20] Many other mathematical functions are of course possible. These two have been discussed as limiting cases of the functions that appear to be most typical of psychological and physiological intensity relationships. Alternative relationships between stimulus intensity and amount of inhibition are not discussed here because as long as inhibition increases with increase in the stimulus, the precise nature of the relationship would not cause the curves to cross each other.

is an artifact of the method of absolute judgment, resulting from the fact that there were no stimuli in the series greater than intensity 4. If this were so, intensity 4 may have taken on an absolute value as the loudest of the series. This possibility can be checked adequately only by an extension of the experiment, but certain facts make it difficult to accept as a complete explanation of this particular fact. It will be remembered that all the S_2's were of this same intensity. In that case it is not immediately evident why only the judgments of S_2 following a first stimulus of intensity 4 should be affected by such an artifact. Furthermore, there was one subject, D, who showed reversal at 285 ms. but not at 177 ms. It would be expected that such an absolute standard would operate throughout a series if it existed at all. It is possible, however, that where S_1 and S_2 are objectively equal, the judgment of S_2 will be more greatly influenced by the absolute standard than in the case where S_2 follows other unequal intensities of S_1. The positive correlation between J_1 and J_2 at the large intensities (to be discussed in the second paper) is consistent with this theory, though it is still difficult to understand why subject D did not show this effect at the short interval.

The second possible explanation of the terminal reversal in the J_2 function lies in some long circuited source of facilitation of J_2 by S_1 which would serve to couteract the increased amount of inhibition which should exist at intensity 4 if the straightline relationship between logarithm of the stimulus intensity and time to maximum inhibition is valid throughout the intensity range. Two different sorts of facilitation might be considered: one appearing only after a period of practice, the other, present from the very beginning of the experiment before practice could have any effect. If the first type of facilitation were present, it would be tempting to seek for a parallel between this and the phenomenon of conditioning in reflexes. We might expect, for example, that after a certain amount of practice, a subject would tend to anticipate the coming of a second stimulus. In this case, it would be plausible to expect the judgments to be reinforced by this anticipa-

tion. This possibility is ruled out at once, however, by the fact that there was no consistent increase in the size of J_2 with practice and that the result reported was present from the very outset. Some subjects showed practice effects, but in at least one of these, concomitant increase in J_1's and J_2's occurred. In the other subjects, J_2 decreased with practice, so that there was not enough consistency to support the conditioning notion.

It is unnecessary to consider what difference there may be in the neural functions accompanying the facilitation which follows on practice as in conditioned responses and that which appears without previous pairing of the stimuli involved. It can only be observed at present that the two kinds of facilitation cannot yet be identified, because of their apparent functional differences, and that they both occur. There is evidence from many studies that, without practice, a shift in the intensity of S_1 may cause inhibition to go over into facilitation. Allen, for example (**1**), has reported that a stimulus applied to the left eye produced 'reflex' effects on the critical flicker frequency of the right eye, and he has shown that the nature of these effects is dependent on the stimulus intensity, so that a wave length which at high intensity produced enhancement, might at a lower intensity elicit depression. Dodge, too, has discussed this point at length (**4**, p. 69), quoting Fröhlich to the effect that "Inhibition is nothing but weak excitation." Humphrey (**10**, Ch. 6) has also considered the relationship of stimulus intensity and resulting inhibitory processes, showing that as a stimulus increases in intensity, there is a shift in the relative values of excitation and inhibition. Such facilitation of J_2 at the large intensity would tend to mask the increase in inhibition.

A third factor which might be acting to counteract the increased inhibition assumed to exist when S_1 is 4 can only be mentioned here, for the data in support of it will be presented and discussed in the second paper of the series. There it will appear that not only is J_2 larger following intensity 4 than after intensity 3, but also that in some subjects under these same conditions, the judgment of the first stimulus of the

pair is smaller than the judgment of the same stimulus alone. This suggests the possibility of an equilibration process between the after-effects of S_1 and S_2.

It is evidently impossible to determine from the present data which of these alternatives is the correct one. The answer must be found in further experiments designed to show whether the J_2 terminal rise indicates the super-position on the functions of Fig. 4 of other factors, or whether the reversal arises out of a difference in thresholds of reflexes and judgments which makes the stimuli of the present experiment functionally unequal for the two types of reaction.

Summary and Conclusions

1. The series of papers of which this is the first has attempted to compare two different behavior functions, as an example of the sort of inductive process that is indispensable to the organization of facts. The experiment has investigated lid reflexes to paired tone stimuli and absolute judgments of these same tones. The four different intensity values taken by the first stimulus of the pair fell approximately into geometrically increasing steps. The second stimulus was of constant intensity. Two time intervals between stimuli were employed, 177 and 285 ms. The 1000 cycle tones were 50 ms. in duration. Data on 5 subjects are reported.

2. As a first step in the analysis of the positive time order error, judgments of and reflexes to the second stimulus (S_2) of a pair have been analyzed. Responses to the first stimulus and the interrelationships between responses will be discussed in another paper.

3. With S_1 and S_2 equal, there is evidence of a positive time error in 3 out of 5 subjects at the 177 ms. interval, and 4 out of 5 at the 285 ms. interval. The error tends to be greater at 177 ms. These are intervals at which Köhler reported no positive time errors.

4. Although judgments of S_2 following all intensities of S_1 showed decrement, the amount varied. The judged intensities of the second stimulus of the pair (J_2) decreased as the intensity of the first stimulus increased. When the

intensity of the first stimulus was greatest and equal to the second, however, the function tended to reverse, showing a decrease in decrement of J_2, although inhibition was still present in significant amounts.

5. The amplitude of the lid reflex (R_2) showed a relationship to the first stimulus similar to that for J_2, except that the function was continuous with no reversal at the largest intensity.

6. The effect of time interval upon both lid reflex and judgment response showed some variation with individuals. In both reactions, however, it was found that the development of inhibition following weak intensities of S_1 tended to be delayed so that it continued to increase beyond 177 ms. and was greater at the longer time interval. At the largest intensities this never occurred and the amount of inhibition was always smaller at the longer intervals.

7. An attempt is made to organize the J_2 and R_2 data under the following generalizations which are supported by experimental observations of spinal reflexes.

a. Inhibitory effects increase with increase in the stimulus intensity.

b. The time to maximum inhibition of the response to the second stimulus varies inversely as the intensity of the first stimulus. Two mathematical functions are discussed here.

c. Duration of inhibitory effects increases with the intensity of the stimulus, and dispersion of inhibitory effects occurs at a slower rate than recruitment.

8. The chief difference in J_2 and R_2 functions is found in the tendency for decrement in J_2 to begin decreasing when S_1 is very intense, whereas for R_2, the decrement is greatest at this intensity. Several possible explanations of this difference are offered.

9. The analysis is presented tentatively as a method of organizing the data which has the advantage of integrating both judgment and lid reflex functions, as well as certain physiological responses already shown to be similarly related to the variables in question. It is not suggested that, if the assumptions prove valid on further experimentation, judgment

responses will be 'explained' by the lid reflex, by spinal reflexes, or vice versa. The reactions will simply tend to fall under the same general principles to the extent that their functions are alike. It is evident that even within the narrow limits of this experiment, certain differences between judgments and reflexes have begun to appear. Other differences will be discussed in the other papers of the series, as well as the relation of the data to theories of the time order error. It will be shown that the formulation offered here needs supplementation in order to care for the facts of comparison judgement observed in the experiment, but that it incorporates a wider range of facts than other available theories.

(Manuscript received June 15, 1939)

REFERENCES

1. ALLEN, J. F., Variation of visual sensory reflex action with intensity of stimulation, *J. Opt. Soc. Amer.*, 1926, **13**, 383.
2. BUJAS, Z., Quelques remarques sur le contraste et l'inhibition a la suite d'excitation gustatives simultanées, *C. r. Soc. Biol.*, 1934, **116**, 1304–1306.
3. CREED, R. S. AND OTHERS, *Reflex Action of the Spinal Cord.* Oxford, Clarendon Press; 1932.
4. DODGE, R., *Conditions and Consequences of Human Variability.* Yale University Press, New Haven; 1931.
5. DODGE, R., *Elementary Conditions of Human Variability.* Columbia University Press, New York; 1927.
6. FORBES, A., The mechanism of reaction, *Handbook of General Experimental Psychology.* Worcester, Clark University Press; 1934, pp. 155–203.
7. HEYMANS, G., *Gesamelte kleinere Schriften Untersuchungen über psychische Hemmung.* Haag, Nijhoff; 1927, Pt. II.
8. HILGARD, E. R., Reinforcement and inhibition of eyelid reflexes, *J. gen. Psychol.*, 1933, **8**, 85–113.
9. HOLOWAY, A. H., GOLDING, L. E., AND ZIGLER, M. H., On the discrimination of minimal differences in weight. IV. Kinæsthetic adaptation for exposure intensity as variant, *J. exp. Psychol.*, 1938, **23**, 536–544.
10. HUMPHREY, G., *The Nature of Learning.* Harcourt, Brace and Co.; 1933.
11. KAROWSKI, T. AND RIGGS, L. A., A note on foveal inhibition by continuous exposure to light, *J. gen. Psychol.*, 1937, **17**, 402–404.
12. KELLOGG, W. N., Measuring auditory intensive thresholds in electrical units, *J. exp. Psychol.*, 1929, **12**, 240–248.
13. KÖHLER, W., Zur Theorie des Sukzessivvergleichs und der Zeitfehler, *Psychol. Forsch.*, 1923, **4**, 113–175.
14. LORENTO DE NÓ, R., Inhibition of Motoneurons. Reprinted from *Problems of Nervous Physiology and Behavior*, 1936.

15. LAUENSTEIN, O., Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler, *Psychol. Forsch.*, 1932, **17**, 130–177.
16. MÜLLER, G. E., *Die Gesichtspunkte und die Tatsachen der Psychophysichen Methodik.* Weisbaden, 1904 (quoted from Needham (**18**)).
17. NEEDHAM, J. G., The effect of the time interval upon the time error at different intensive levels, *J. exp. Psychol.*, 1935, **18**, 530–543.
18. NEEDHAM, J. G., The time error in comparison judgments, *Psychol. Bull.*, 1934, **31**, 229–243.
19. PEAK, H., Inhibition as a function of stimulus intensity, *Psychol. Monog.*, 1936, **47**, 135–147.
20. PEAK, H., Modification of the lid reflex by voluntarily induced sets, *Psychol. Monog.*, 1931, **42**, No. 188.
21. PEAK, H., The time error as a function of judgment of the first stimulus of a pair, *This Journal* (in press).
22. PEAK, H., Graded inhibition at sub-threshold intensities (In preparation).
23. PRATT, C. C., Time errors in the method of single stimuli, *J. exp. Psychol.*, 1933, **16**, 798–814.
24. PRATT, C. C., Time error in psychophysical judgments, *Amer. J. Psychol.*, 1933, **45**, 292–297.
25. SHERRINGTON, C. S., Remarks on some aspects of reflex inhibition, *Proc. Roy. Soc.*, London, 1925, **97B**, 519–545.
26. STEVENS, S. S. AND DAVIS, H., *Hearing.* John Wiley, New York; 1938.
27. VOLKMANN, J., The method of single stimuli, *Amer. J. Psychol.*, 1932, **44**, 808–809.

GOAL GRADIENT, ANTICIPATION, AND PERSEVERATION IN COMPOUND TRIAL-AND-ERROR LEARNING [1]

BY CHESTER JAMES HILL

Institute of Human Relations, Yale University

Introduction

The goal gradient hypothesis of Hull (**8**) has been the instigator of considerable experimental work. Most of this work was devoted to supporting or refuting the hypothesis by showing the presence or absence of the backward elimination of errors in the ordinary types of mazes, a theorem deduced from the hypothesis. A considerable number of these experiments have shown that this error elimination is complicated by several specific factors in the maze, among which are centrifugal swing (**1**), goal orientation (**4**), and preponderance of correct turns in one direction (**15**). Probably the two most important factors influencing the error elimination are anticipation [2] (**3, 17**) and the forward spread of reinforcement (**13**).

Since the ordinary maze contains only two different kinds of turns—right and left—it is impossible to tell whether errors at a specific point are anticipations of the last correct turn or some other similar correct turn intervening between the error point and the goal, or even a possible perseveration from a preceding choice. In order to study these factors, a maze was used which would avoid differential effects from centrifugal swing, goal orientation, and preponderance of similar turns. It is possible, as Spragg (**17**) points out, that the anti-

[1] This experiment is part of the integrated research program of the Institute of Human Relations. The author is indebted to Dr. Clark L. Hull for many kinds of aid.

[2] Anticipation is defined by Lumley (**10**) and Spragg (**17**). It refers to an experimenter's classification of responses which appear inappropriate at a certain point in behavior but which would be appropriate if occurring later. Perseveration may be defined conversely.

cipatory errors are a function of the goal gradient itself; however, these errors of anticipation and perseveration are interesting enough in themselves to warrant study.

Apparatus and Procedure

The apparatus consisted of a maze and waxed-paper polygraph as shown in the lower diagram of Fig. 1.

The maze, a modification of one section of the type used by Muenzinger, Dove, and Bernstone (**13**), was a straight runway 12 feet long, 1 foot wide, and $6\frac{1}{2}$ inches high. It was constructed of wood and painted black to reduce visual cues to a minimum. The maze was covered by 5 wooden frames of $\frac{1}{4}$-inch wire screen, each frame being hinged to the maze to facilitate cleaning. Four choice points, each of which had 4 short alleys pointing in the direction of the goal, divided the maze into 5 sections 28.8 inches in length. Each of these sections had, midway between choice points, a partition with a small passageway in its center which forced the animal to make a choice from the same position at each choice point. Identical partitions were placed midway between the starting point and first choice point, as well as between last choice point and goal.

The upper diagram of Fig. 1 presents a close-up of a choice point. In each of the short alleys was a valve of light plywood, hinged at the top and sloped at such an angle (60°) that it was easily lifted by pressure of the rat's head when applied in the correct, *i.e.* goalward, direction. After the rat passed through a valve, the valve fell back to a closed position through its own weight and fitted snugly to the floor of the maze. Neither pressure nor clawing on the other side served to open the valve, so that retracing was an impossibility.

Since only one valve which could be opened was necessary at any choice point and since this was different at each choice point and different for each animal, a readily changeable method of blocking the other 3 valves at a choice point was desirable. This was accomplished by a bar with a slot (*B* in Fig. 1), which acted as a doorstop by blocking the heavy brass strips extending vertically upwards from the top of

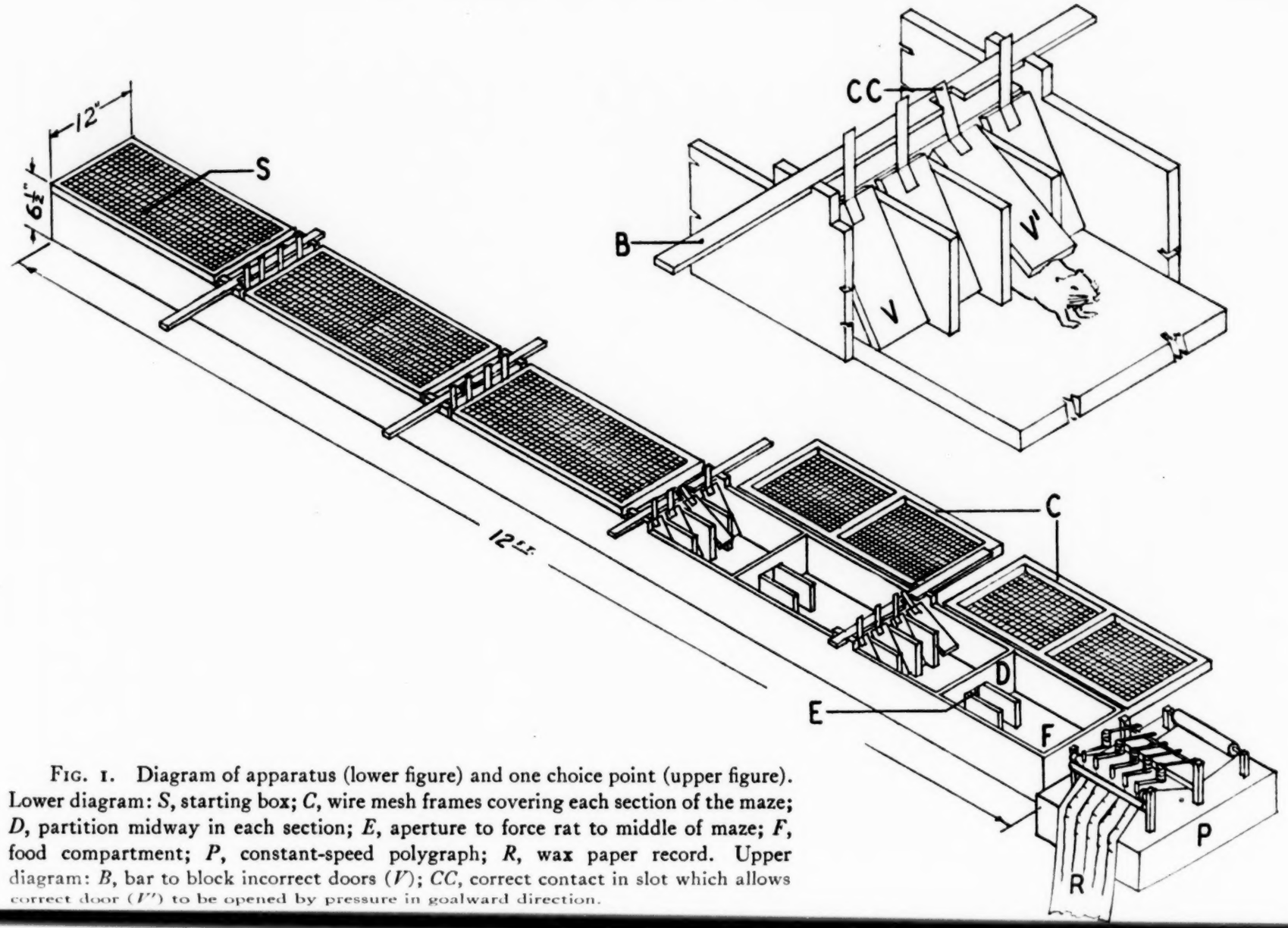


FIG. 1. Diagram of apparatus (lower figure) and one choice point (upper figure). Lower diagram: *S*, starting box; *C*, wire mesh frames covering each section of the maze; *D*, partition midway in each section; *E*, aperture to force rat to middle of maze; *F*, food compartment; *P*, constant-speed polygraph; *R*, wax paper record. Upper diagram: *B*, bar to block incorrect doors (*V*); *CC*, correct contact in slot which allows correct door (*V'*) to be opened by pressure in goalward direction.

each valve. The bar, since it was inserted in grooved runways on the sides of the box, could be moved back and forth across the box by the experimenter so that the slot could be placed at any valve desired. Whenever this valve was tried by the rat, the brass strip could move back into the slot of the bar, thereby allowing the rat to open the valve and pass through.

Errors, correct choices, and times were recorded electrically. The brass strips of those valves in a similar position at every choice point were wired in parallel to a signal marker. Thus, there were 4 signal markers for the 16 doors: all doors on the left side at each choice point were wired to one signal marker; all doors one from the left, to the second signal marker; and so on. The circuit for each of these markers was closed by the contact of the brass block on the valve and a strip of brass on the side of the bar.

In addition, the correct door at each choice point recorded on a fifth marker; for when the brass rod moved back into the slot of the bar it closed a spring-brass contact and completed the circuit of the fifth signal marker as well as its own identifying circuit. Inasmuch as the open valve at any choice point ruled out the possibility of the similar valve being open at any other choice point, there was no ambiguity in reading the records. A bell button, wired in the fifth marker circuit, was located on the outside of the box at the starting point. This was pushed by the experimenter when the rat was placed at the beginning of the maze and was useful in measuring starting latency.

The recording was made on waxed polygraph paper, driven at the rate of 1/10 inch per second by a constant-speed telechron motor. Time records could be obtained with a minimum of effort by using a conveniently graduated scale, placed directly on the record. A somewhat reduced sample record is shown in Fig. 2.

The subjects consisted of 72 male albino rats from the Brooklyn Breeding and Laboratory Institute, and had no previous experience in experimental situations. They were fed on the regular laboratory diagram: *B*, bar to block incorrect doors (*I'*); *CC*, correct contact in slot which allows correct door (*I''*) to be opened by pressure in goalward direction.

½ hour after each experimental period, so that motivation could be assumed to be quite constant throughout the 12 days of the experiment.

In the preliminary training the animals were taught to go through the valves. This training period lasted 2 days. On the first day each animal was placed in the compartment before the last choice point and was allowed to squeeze under any one of the valves, all of which were partly open, and obtain food in the food compartment. After the rat had eaten the food it was again returned to the compartment just

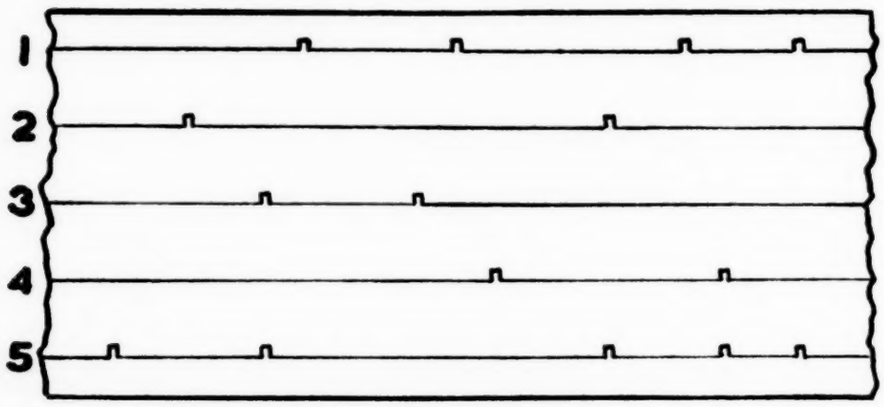


FIG. 2. A transcript polygraph record recording errors, successes, and times of 1 trial (Scale is 1/3). Lines 1 to 4 record attempts at valves 1 to 4, respectively. Line 5 records the start of the trial or success at any choice point. The correct valve order in the above figure is 3–2–4–1. The record is read left-to-right, thus: The rat is placed in the starting box (5), tries valve 2 (2), passes through valve 3 (3 and 5), tries valve 1 (1), tries valve 3 (3), tries valve 1 (1), tries valve 4 (4), passes through valve 2 (2 and 5), tries valve 1 (1), passes through valve 4 (4 and 5), and passes through valve 1 (1 and 5).

before the last choice point and allowed to obtain food a second time in the same manner as before. After 5 trials of this sort, 5 additional trials were given, this time with all valves closed but none barred. The rat was then returned to its living cage.

On the second training day the rat was placed in the starting compartment and allowed to make 10 rewarded runs in the maze with all valves shut but none barred. The experimenter kept a record of the valve used at each choice point, in order to discover whether this method of training was

establishing decided preferences in the group for any one door. After the tenth trial the animal was returned to its living cage.

On the third day the experiment proper began. On this day the rat was placed in the starting box and was presented with a definite arrangement of unbarred valves. No valve was the unbarred one at more than one choice point, and this pattern was kept constant throughout the experiment for each animal. The rat was left in the maze until it found its way through the 4 choice points and had eaten the food pellet at the goal. It was immediately given 4 more trials and then returned to its living cage where it was fed for ½ hour. This procedure was repeated at the same time on the next day and continued for 10 days altogether, making a total of 50 trials.

In order to eliminate any effect of the particular pattern of correct doors, all permutations of doors were used—24 in all—and 3 animals were run with each permutation.

Results

Table I presents the average number of times each valve

TABLE I

Average Number of Passages for Each Valve During 10 Rewarded Training Trials

Choice Point	Valve			
	1	2	3	4
1	3.06	2.52	2.17	2.25
2	2.42	3.52	2.19	1.87
3	2.38	3.35	2.75	1.52
4	1.81	3.94	3.06	1.19
Total	9.67	13.33	10.17	6.83

was chosen by the rats during the 10 training trials of the second day. The first choice point has a distribution of passages through each of the 4 valves very close to that which is to be expected on a purely chance basis, *i.e.* 2.5 passages per valve. However, as the goal is approached, the middle valves have an increasing number of passages, while the valves

adjacent to the walls (1 and 4) have a decreasing frequency of choice. These trends seem quite reliable, as they are consistent except for one small deviation, the second valve at the third choice point displaying a slightly lower frequency of passage than the corresponding valve at the second choice point.

Since the middle valves constitute paths to the goal which are shorter than the paths of the valves adjacent to the walls, this factor in itself would be sufficient cause for the preponderance of middle-valve choices. This is in agreement with the results of DeCamp (**5**) and Yoshioka (**20**). In addition to this the *increasing* frequency of short-path responses as the goal is approached suggests that the short path is receiving differential reinforcement according to its distance from the goal; that is, the short paths closest to the goal are most strongly reinforced. This has been shown experimentally by Yoshioka (**20**) and is a corrollary deduced from the goal gradient hypothesis by Hull (**8**). This corollary states that the smaller the relative difference in distance to the goal between two alternative paths, the more difficult will be the discrimination. Though the difference between short paths and long paths at any choice point is constant, the relative difference increases as the goal is approached so that more short-path runs should be taken at choice points closer to the goal.

From the above it is clear that during training the rats learned to choose the center valves. However, it seems justifiable to assume that the use of all possible permutations of valves would balance out this factor.

Table II shows the errors made at each choice point for sets of 10 trials. It was not uncommon for a rat to persist in its attempt to open a door which was barred so that consecutive pushes were sometimes as frequent as several per second. It was not deemed advisable to count these as discrete errors so, unless a period of 10 seconds with no further attempts or a try on another valve intervened, these pushes were recorded as 1 error. The interval of 10 seconds

as a criterion of discrete errors at one valve was believed to be adequate to indicate unrelated errors at that valve.

From examination of Table II one can see that in the early trials (the first 10) a single gradient of errors does exist. Choice point 4, the closest to the goal, has the fewest errors, and each choice point farther removed has an increasing number of incorrect attempts. This is exactly the behavior demanded by the goal gradient *if uncomplicated by anticipatory errors and their extinction effects.* This situation is of short duration, however, for in the next set of 10 trials a compound, or double, gradient arises which becomes more and more marked as learning progresses. The first and last choice points display considerably fewer errors than the 2

TABLE II

TOTAL ERRORS AT EACH CHOICE POINT

Sets of 10 Trials	Choice Point			
	I	II	III	IV
1	1367	1168	1131	954
2	614	684	694	342
3	404	482	485	115
4	299	338	303	72
5	202	263	251	33

middle choice points, and the general picture is that displayed by human maze and rote learning (**14, 19**)—that is, a forward-backward elimination of errors.

The compound gradient [3] of errors does not arise from an increase in errors at the middle 2 choice points, but rather from the very slow elimination of errors at these places relative to the speed of error elimination at the first and last choice points. Errors are reduced in amount 55 percent at the first choice point and 64 percent at the last from the first 10 to the second 10 trials; the same figures for the second and third choice points are 41 percent and 39 percent, respectively.

In order to analyze the factors at work in producing the above effect, the data were sorted for the different types of

[3] The term *double*, or *compound*, gradient refers to a series of values changing gradually, first in one direction, then in another.

errors and are presented in Table III. This table contains the errors segregated according to degree of anticipation or perseveration of the correct selection at each choice point upon every other choice point. These data indicate that anticipatory errors drop out very slowly, particularly anticipatory errors one removed from the choice point where they would be correct. Thus, the errors at the third choice point

TABLE III

SUCCESSES AND DIFFERENTIAL ERRORS *

$N = 72$

Choice Point	Origin	Sets of 10 Trials					Total
		I	II	III	IV	V	
1	1	*233*	*385*	*484*	*542*	*612*	*2256*
	2	424	218	175	140	118	1075
	3	454	198	108	91	53	904
	4	489	198	121	68	31	907
2	2	*235*	*364*	*426*	*487*	*543*	*2055*
	1	408	183	94	64	54	803
	3	388	272	221	176	161	1218
	4	372	229	167	98	48	914
3	3	*236*	*328*	*411*	*447*	*527*	*1949*
	1	344	167	100	85	33	729
	2	385	187	137	98	50	857
	4	402	340	248	220	168	1378
4	4	*281*	*492*	*639*	*653*	*693*	*2758*
	1	296	99	33	15	7	450
	2	305	96	37	27	11	476
	3	353	147	45	30	15	590

* The first column indicates the choice point at which the error occurred; the second column, the choice point where the valve selection would have been a correct choice. Thus, the first row is read: successes at choice point 1; the second row is read: anticipations of choice point 2 at choice point 1; and so on. The italicized figures represent successes.

anticipating the fourth choice point, those at the second choice point anticipating the third, and those at the first choice point anticipating the second are considerably more frequent than the other errors consisting of remote anticipations and perseverations. The fact that these anticipations, occurring just one point too early in the behavior sequence,

increase at each choice point closer to the goal seems an indirect verification of the goal gradient hypothesis, although the actual error gradient does not possess a shape similar to that which might be expected from the goal gradient hypothesis alone.[4]

Another fact worthy of notice is that anticipations of any choice point become increasingly frequent as that choice point is approached. For example, the total anticipations of choice point 4 at choice points 1 to 3, respectively, are: 907, 914, and 1378. This statement holds true not only for the totals but also for *each* choice point for *every* set of 10 trials except the first 10. As indicated above, the generalization may be formulated that at any choice point contiguous anticipations are more frequent than remote, but they also become relatively greater in number as learning progresses. Thus, at choice point 1 (in the second set of 10 trials), the anticipations of choice point 3 are 91 percent as frequent as anticipations of choice point 2, but in the last set of 10 trials the same figure is only 45 percent.

Perseverative errors drop out quickly. By chance, they should be present to the same amount as anticipations, for there is equal opportunity for both types; but even in the first 10 trials they are only 45 percent of the total errors, and by the last 10 trials they are a mere 32 percent.

As might be expected from what is known of the duration of perseverative effects, correct choices from the choice point *immediately* preceding any choice point perseverate more strongly to the latter than they do to the choice points 2 or 3 removed. The correct valve at choice point 1 displays a total of 803 perseverative errors at the second set of valves, 729 at the third, and 450 at the fourth. Intimately related to this is the amount of perseveration at each choice point as a function of its degree of remoteness from every other choice point. At choice point 4, the correct valve from the choice point farthest removed (1) perseverates 450 times, the correct valve from choice point 2, 476 times, and the correct

[4] Hull and Perkins, using a mathematical approach, have analyzed these data in respect to the goal gradient hypothesis.

valve from the immediately preceding choice point (3), 590 times. From these data it appears that perseverative errors at any choice point are more frequent for those valves correct at closer choice points than for those valves correct at more remote choice points.

Data on the successes at each choice point are also contained in Table III. By successes, of course, is meant the selection of the correct valve at any choice point with no preceding error or errors at that same choice point. The successes take a course very similar to that of the errors—that is, there is a tendency for a single gradient at first, with more successes as the goal is approached, and then, by the second set of 10 trials, a definite superiority of the beginning and end choice points, which continues until experimentation ceases. That the problem is not insoluble for rats is shown quite clearly by the consistent trend at every point toward more successes per 10 trials. In fact, the last choice point in the last 10 trials has a total of 693 out of a possible 720, or 96 percent correct choices; while even the third choice point —that most poorly learned—has 73 percent correct choices.

Since generalization of inhibition or frustration is a common occurrence among responses associated with similar stimuli, as has been shown by Ellson (7), Mitrano (11), and Youtz (21), it was thought that possibly the error gradients would show different relationships if only the first selection at any choice point were used than if all choices were taken into consideration. By the former method, repeated frustrations at a barred valve at any choice point might affect other choices there unequally. This method of selecting only the first choice should avoid any differential generalization of frustration among the selections at that choice point.

In Table IV are the results of this computation. On the whole, the error elimination is very similar to that obtained from the method using all errors. The only discrepancy of any amount is in the anticipatory errors of the fourth, the last choice point, at choice point 3. In this method using only first selections the errors of this type do not decrease during early learning but actually *increase* from 193 in the

first 10 trials to 239 in the second 10 trials—an increase of errors of 42 percent during learning. This must mean that although the rat is now anticipating the last choice point more on the first selection at the third choice point, it is not repeating this selection frequently as it did in the first 10 trials; thus, in the first 10 trials, the first-selection errors anticipating choice point 4 are only 48 percent of the total

TABLE IV

DIFFERENTIAL FIRST-CHOICE ERRORS AND SUCCESSES *

$N = 72$

Choice Point	Origin	Sets of 10 Trials					Total
		I	II	III	IV	V	
1	1	*233*	*385*	*484*	*542*	*612*	*2256*
	2	175	123	102	98	59	557
	3	146	107	66	48	30	397
	4	163	94	64	31	19	371
2	2	*235*	*364*	*426*	*487*	*543*	*2055*
	1	159	65	40	28	15	307
	3	156	140	144	126	127	693
	4	168	147	97	71	34	517
3	3	*236*	*328*	*411*	*447*	*527*	*1949*
	1	129	63	31	27	12	262
	2	160	90	63	38	25	376
	4	193	239	211	206	152	1001
4	4	*281*	*492*	*639*	*653*	*693*	*2758*
	1	132	56	25	14	6	233
	2	142	57	21	25	8	253
	3	151	103	35	26	13	328

* Read as Table III. This table consists of only the first selection made at each choice point.

errors (193 out of 402) anticipating choice point 4, while in trials 11 to 20 they constitute 70 percent (239 out of 340).

The forward-backward elimination of errors resembles to a great extent that bidirectional gradient of errors found by Muenzinger, Dove, and Bernstone (**13**). They used a maze with 4 units, each similar to the one in this experiment, placed end to end in the form of a square so that the animal ran from goal to goal. Those experimenters attribute their results to the Thorndike effect—reward spreading forward

and backward. Upon casual observation of the results, the same factor might be thought to be operative here. Since the animals were returned to the start immediately after eating the food at the goal, the start and first choice point, being so close to the goal in the animals' behavior sequence, might receive enough reinforcement from the forward spread of reward so that the start and first choice point would show fewer errors. This is quite conceivable in the light of the work of Brandt (**2**), showing that the spread of reward from a successful response is not a function of the time between units but of the number of units.

This view seems untenable in this experiment for several reasons. First, the single gradient is present early in training. There seems to be no reason why the double gradient should not be existent early in training if it is the result of this forward spread of reinforcement. Thorndike (**18**) and Muenzinger and Dove (**12**) have shown this double gradient to exist after the first trial with human beings. In addition, on the first trial of each day the start and first choice point are removed both in time and activity from the reward effects of the last trial of the preceding day, yet the error gradient of the first trial each day is very consistent with the error gradient of all trials. The total errors of first trials each day in trials 1 to 10 are, for choice points 1 to 4, respectively: 307, 260, 226, and 191. This is a single gradient. However, the same figures for trials 41 to 50 are: 44, 58, 50, and 9. A comparison of these data with the data for all trials of the day shows that the gradients are similar. Probably the most convincing evidence against the view that the bidirectional gradient observed here is a function of bidirectional reinforcement is the data from Table III, which shows that the bidirectional gradient arises from a failure to eliminate anticipatory errors adequately at the middle choice points.

Hull (**9**) has shown that in the early trials on a straight runway a rat's speed of locomotion increases as the goal is approached. This speed gradient flattened with further learning. Drew (**6**), using a circular runway, has pointed out recently that this gradient is a function of the massing of

trials. He found that the first trial of the day displayed no gradient of speed but that the successive trials presented gradients with increasing slopes. This was not caused by increased speed in the last sections of the maze but, rather, by a retardation in the early sections.

Table V presents the average times consumed by the rats in running the different sections of the maze. Only perfect runs were used in these data because errors would certainly exert different influences on the times spent in the various sections. These results are presented graphically in Fig. 3. Quite a perfect gradient is displayed by the times of all perfect trials in trials 1 to 20; however, it must be kept in mind that the first choice point of the graph, *i.e.* start to the first choice point, is not strictly comparable to the other points as it measures the time elapsing from the insertion of

TABLE V

TIME IN SECONDS BETWEEN CHOICE POINTS OF PERFECT TRIALS

Trial	Perfect	Start—1	1–2	2–3	3–4
1–20	All	4.1	2.5	2.0	1.9
	First of day	2.4	2.5	1.8	2.4
	Last of day	4.8	2.2	2.7	1.8
31–50	All	3.3	2.0	1.9	1.6

the animal into the maze to the transit of the animal through the first choice point. A segregation of the perfect runs of the first 20 trials, which were last runs of the day, shows an *increased* amount of time taken by the rat in transversing the first section of the maze, while a similar separation of those runs which were *first* of the day shows a considerably *decreased* amount of time in the first section of the maze. These data support the work of Drew (**6**), although the effect here seems limited to starting latency, as the times for the other sections of the maze are quite irregular. It is interesting to note that these latter 2 curves are inversely related, though the inversion may occur only by chance.

The average for each section of the maze for the perfect trials in trials 31 to 50 is also found in Fig. 3. The gradient

is still preserved, although speed of locomotion in all sections has increased. This increase in speed is greatest for the section start to choice point, which may be a forerunner of

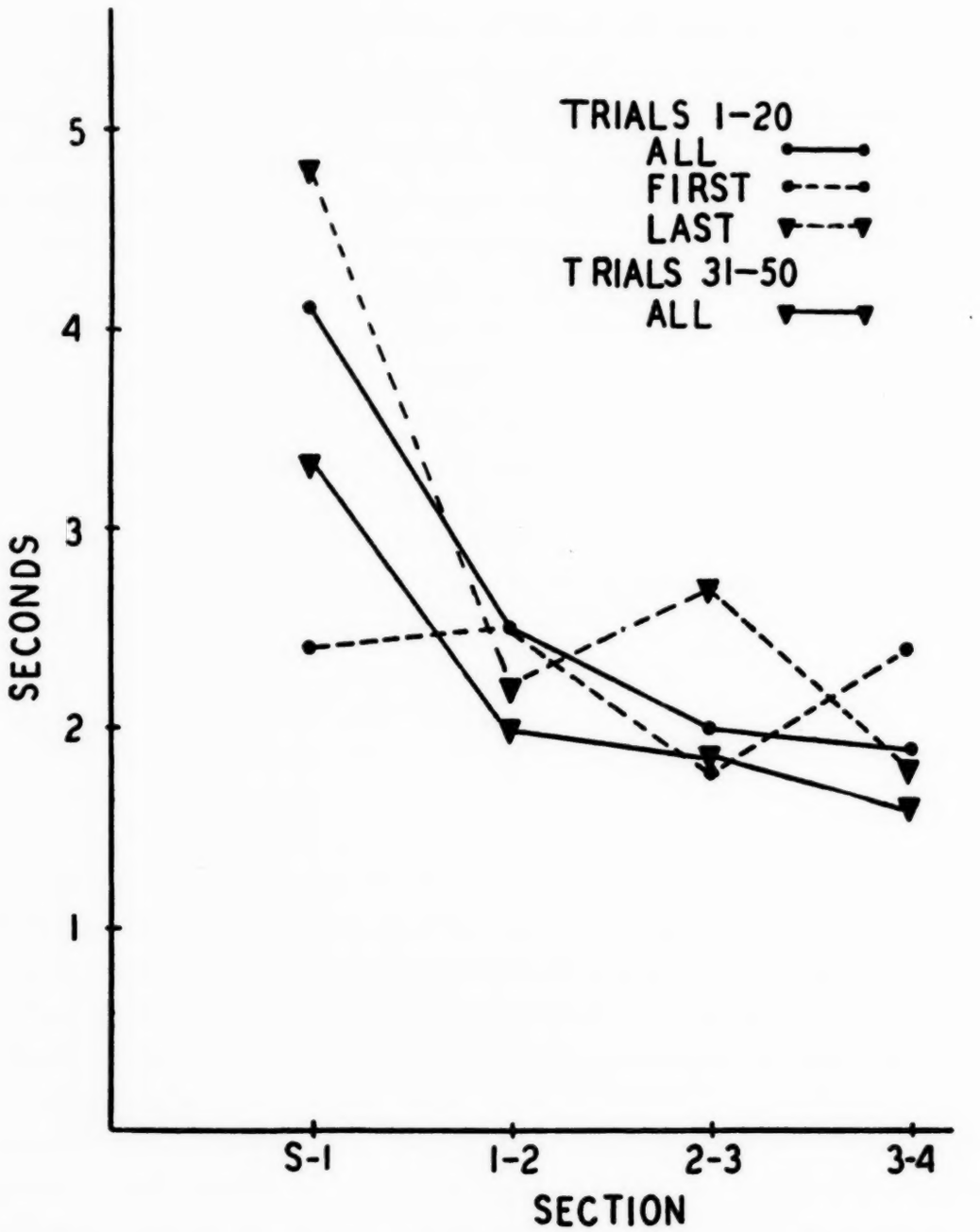


FIG. 3. Mean time per section of perfect trials. *First* and *last* signify first and last of the day's trials.

the flattened gradient in advanced training discovered by Hull (9). Of course, the fact that the whole curve is lowered is indicative of future flattening, as a definite physiological

limit would soon be reached in those sections just preceding the goal—a condition necessitating a curve asymptotic to the base line.

To determine the effects of the removal of reward (extinction), one group of animals having a complete set of the different permutations in learning ran 10 unrewarded trials the day after the last learning trials had occurred. The first selections at each choice point are given in Table VI.

The first noteworthy fact is that correct choices for each choice point are less frequent, relative to the number of errors, than those same choices in the last day of learning (Table IV). On that day the percentages of errors for choice points 1 to 4,

TABLE VI

DIFFERENTIAL FIRST-CHOICE ERRORS AND SUCCESSES DURING EXTINCTION *
$N = 24$

Choice Point	Origin	Choices	Choice Point	Origin	Choices
1	1	*150*	3	3	*114*
	2	38		1	36
	3	40		2	43
	4	12		4	47
2	2	*126*	4	4	*158*
	1	32		1	18
	3	38		2	30
	4	44		3	34

* Read as Table IV.

respectively, are: 85, 75, 73, and 96. The corresponding percentages for the extinction trials are: 63, 53, 48, and 66. From these results, the general conclusion can be formulated that the maze habit is undergoing disruption. Moreover, there is some indication that the disruption is greatest at the last choice point and becomes gradually less at each point removed from the goal, since the number of successes relative to errors at each choice point has decreased 30 percent at the last point, 25 percent at the third, and only 22 percent at both the second and first.

Another point brought out by these data is that the anticipatory errors no longer constitute a preponderance of

errors, as is the case during the last rewarded trials. Although still more frequent than perseverative errors, this difference is very slight. Whether this effect can be attributed to the disorganized, frantic behavior during extinction, or whether it is a direct inhibitory process extending backward from the goal is impossible to tell by observation.

The times in seconds for successful runs during the extinction trials are as follows: start to choice point 1—7.3; choice point 1 to 2—3.6; choice point 2 to 3—3.3; and choice point 3 to 4—2.6. Removal of reward leads to a steeper gradient of running time and increased time in each section of the maze, as Hull (**9**) has shown.

Discussion

It has been shown that early in learning a gradient of errors exists, which is in conformance with the predictions from the goal gradient hypothesis, but this gradient soon is replaced by a compound gradient which arises from the failure to eliminate anticipatory errors, particularly in the middle sections of the maze. These numerous errors are indicative of a high degree of generalization from one choice point to the next. This is to be expected, as the choice points are identical, and on an external stimulus basis small differential reinforcement effects would be possible. Reinforcement of a correct response would reinforce the tendency to make this response at other choice points. In a runway where differential stimuli are at a minimum, it is conceivable that generalization may exist to such an extent that sections distant from the goal are as strongly associated with reward as are sections near the goal. This may account for the rather flat speed graph obtained by Drew (**6**) on the first trials of each day.[5]

In addition to the external stimuli there were also stimuli arising and persisting within the organism, and at any given moment it is possible that an accumulation of these stimulus traces exists in the organism. Early in the maze stimulus differences between the first and second sections would be

[5] These and related ideas are included in the memoranda of Clark L. Hull's Graduate Seminar, Yale University, 1937–1938.

large and generalization from the first section to the second, small. However, after several sections have been passed through, the action and external stimulation have been constant, and the system of compound stimulus traces is changing slowly so that there is much generalization and frequent errors. On this theory one could predict that it would be difficult for rats to learn a maze of 7 correct right turns, followed by 1 left, as Spragg (**17**) found, but easy to learn a maze of 1 correct right, followed by 7 lefts, as Snygg (**16**) has shown. In the former, generalization would interfere with learning; in the latter, it should facilitate learning.

Summary and Conclusions

In a maze with 4 four-valve choice points an equal distance apart, 72 albino rats were given 10 rewarded training trials with all valves open. Three rats were then run in each possible permutation of open valves 5 trials a day for 10 days; at the end of a day's experimentation, each rat was fed for $\frac{1}{2}$ hour. After these 10 days, 24 rats—1 in each possible permutation—ran 10 unrewarded trials. Errors, successes, and times were recorded electrically by a waxed-paper polygraph.

1. During training (all valves open) the rats take the shorter paths (middle doors) with increasing frequency as the goal is approached.

2. A single gradient of errors is displayed early in learning, the errors decreasing at each point as the goal is approached; by the second set of 10 trials the bidirectional gradient makes its appearance, with fewer errors on the first and last choice points than on the middle 2 choice points.

3. The bidirectional gradient appears to be caused by failure of anticipatory errors to be rapidly eliminated. Anticipations one removed become increasingly frequent at each choice point toward the goal.

4. Successes at each point present gradients similar to the error gradients. Despite the large number of anticipations, the maze appears to be soluble, as successes were still increasing when experimentation ceased.

5. An analysis of only first-selection errors reveals a

gradient similar to the gradient obtained when all errors were computed.

6. A speed-of-locomotion gradient is revealed in all perfect trials of the first 20. Those successes in trials 1 to 20 which were first of the day display decreased time in the section, start to first choice point; those same trials which were last of the day show increased time in that section. No evidence exists for the gradient in the latter two analyses.

7. The speed-of-locomotion gradient of all perfect trials in the last 20 is similar to, but lower than, the gradient of the perfect trials in the first 20.

8. That removal of reward exerts a disruptive effect on the maze habit is shown by increased errors and decreased successes. The effect seems to be greater at each choice point closer to the goal. Under these conditions anticipatory errors decrease in frequency relative to perseverative errors.

Bibliography

1. Ballachey, E. L., and Buel, J., Centrifugal swing as a determiner of choice-point behavior in the maze running of the white rat, *J. comp. Psychol.*, 1934, **17**, 201–223.
2. Brandt, H., The spread of the influence of reward to bonds remote in sequence and time, *Arch. Psychol., N. Y.*, 1935, **27**, No. 180.
3. Buel, J., and Ballachey, E. L., Choice-point expectancy in the maze running of the rat, *J. genet. Psychol.*, 1934, **45**, 145–168.
4. Dashiell, J. F., Direction orientation in maze running by the white rat, *Comp. Psychol. Monogr.*, 1930, **7**, No. 2.
5. DeCamp, J. E., Relative distance as a factor in the white rat's selection of a path, *Psychobiol.*, 1920, **2**, 245–253.
6. Drew, G. C., The speed-of-locomotion gradient and its relation to the goal gradient, *J. comp. Psychol.*, 1939, **27**, 333–372.
7. Ellson, D. G., Quantitative studies of the interaction of simple habits. 1. Recovery from specific and generalized effects of extinction, *J. exp. Psychol.*, 1938, **23**, 339–358.
8. Hull, C. L., The goal gradient hypothesis and maze learning, *Psychol. Rev.*, 1932, **39**, 25–43.
9. Hull, C. L., The rat's speed-of-locomotion gradient in the approach to food, *J. comp. Psychol.*, 1934, **17**, 393–422.
10. Lumley, F. H., Anticipation of correct responses as a source of error in the learning of serial responses, *J. exp. Psychol.*, 1932, **15**, 195–205.
11. Mitrano, A. J., A preliminary investigation of human motivation by conditioned-reaction methods. Ph.D. Thesis, Yale University, 1937.
12. Muenzinger, K. F., and Dove, C. C., Serial learning: I. Gradients of uniformity and variability produced by success and failure of single responses, *J. gen. Psychol.*, 1937, **16**, 403–413.

13. MUENZINGER, K. F., DOVE C. C., AND BERNSTONE, R. H., Serial learning: II. The bi-directional goal gradient in the endless maze, *J. genet. Psychol.*, 1937, **50**, 229–241.
14. ROBINSON, E. S., AND BROWN, M. A., Effect of serial position on memorization, *Amer. J. Psychol.*, 1926, **37**, 538–552.
15. RUCH, F. L., Goal direction orientation, generalized training habit, and goal gradient as factors in maze learning in the rat, *J. comp. Psychol.*, 1934, **17**, 225–232.
16. SNYGG, D., The relative difficulty of mechanically equivalent tasks: II. Animal learning, *J. genet. Psychol.*, 1935, **47**, 321–336.
17. SPRAGG, S. D. S., Anticipation as a factor in maze errors, *J. comp. Psychol.*, 1933, **15**, 313–329.
18. THORNDIKE, E. L., An experimental study of rewards, *Teach. Coll. Contr. Educ.*, 1933, No. 580.
19. WARDEN, C. J., Primacy and recency in maze learning, *J. exp. Psychol.*, 1924, **7**, 98–116.
20. YOSHIOKA, J. G., Weber's law in the discrimination of maze distance by the white rat, *Univ. Calif. Publ. Psychol.*, 1929, **4**, 155–184.
21. YOUTZ, R. E. P., The weakening of one Thorndikian response following the extinction of another, *J. exp. Psychol.*, 1939, **24**, 294–304.

SPONTANEOUS RECOVERY OF THE GALVANIC SKIN RESPONSE AS A FUNCTION OF THE RECOVERY INTERVAL[1]

BY DOUGLAS G. ELLSON

Stanford University

In a previous investigation (**4**) using the bar-pressing response in white rats, the writer obtained curves of spontaneous recovery as a function of the interval following extinction. While theoretical differences between the bar-pressing behavior and the so-called 'typical' conditioned response have been pointed out (**5**, **17**), experimental work with the bar-pressing response (**4**, **18**, **19**, **20**) has shown numerous relationships analogous to those found with the conditioned response. Recent experiments, however (**2**, **3**, **15**), have indicated that the conditioning of human subjects may be influenced by factors which are presumably not present in animal experiments, namely instructions or a verbal evaluation of the situation. The present investigation was performed in order to make a further comparison of the two types of experiment. It utilized the conditioned galvanic skin response in human subjects under conditions closely paralleling those used in the previous animal experiment (**4**). Its plan was to plot the curve of spontaneous recovery as a function of the interval following extinction. Two measures of the strength of the response tendency were used: (i) mean amplitude of the first two unreinforced responses in an extinction series, and (ii) the number of unreinforced presentations

[1] This paper, together with 'Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction' (**4**), formed a part of a dissertation presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Yale University. The experiment is a part of the co-ordinated research program of the Institute of Human Relations. It was carried out in the research unit of Professor Clark L. Hull, to whom the writer is greatly indebted for suggestions and criticism.

of the conditioned stimulus required for extinction to a criterion.

Apparatus

The apparatus has been described in detail by Hovland (7), with one exception: the response in the present experiment was recorded in millivolts by means of a Burr-Lane-Nims microvoltmeter (1) and a General Electric photo-electric recording milliammeter. The subject sat in a comfortable chair in a sound-shielded room. The tonal conditioned stimuli were delivered through earphones, and the shock through copper electrodes on the subject's left wrist. The response was obtained through polished silver (quarter-dollar) electrodes from the back and palm of the subject's right hand. The timing of stimulus presentation was controlled by a Dodge pendulum.

Subjects and Procedure

The procedure, like the apparatus employed in this investigation, was patterned on that used by Hovland, with certain modifications. The galvanic skin response was conditioned to two stimuli. One of these stimuli was then presented without reinforcement until the average magnitude of the response in three successive presentations was equal to or less than one third of its magnitude at the beginning of extinction. After reaching this criterion, the subject was given a 'rest period' after which the response was again extinguished to the same criterion. The duration of the 'rest period' or recovery interval, was five minutes, twenty minutes one hour, or three hours. This parallels the procedure used with rats in the previous experiment (4). Two conditioned stimuli, while not necessary for the experiment as it was performed, were used in order to make it possible to run an experiment on recovery from generalized extinction effects similar to that described as Experiment II in the previous paper. The second experiment, however, has not been performed.

The two tones used as conditioned stimuli were selected from those used by Hovland in his studies of the generalization of the conditioned galvanic skin response (7, 8, 9, 10). Tone *A* had a frequency of 153 d.v., and tone *B*, 1967 d.v. These tones were approximately 75 j.n.d.'s apart. The tones were equated for loudness to a 1000 d.v. tone 40 db. above threshold. A detailed account of the standardization of these stimuli is given by Hovland in the first of his series (7).

The successive steps of the procedure of the present experiment were as follows:

I. The subject was seated in a Morris chair in the sound-shielded room and asked if he knew the nature of the galvanic skin response. If he did not, he was given a brief description before the recording electrodes were attached, and was told that he would not receive a shock through these electrodes. The door to the shielded room was then closed. After a two-minute wait, the subject's response to each of the two tones was tested twice in counterbalanced ABBA order.

II. The experimenter reëntered the subject's room, fastened the shock electrodes on the subject's wrist, and adjusted the shock. The subject was instructed as follows: "The shock will be built up from zero to the point where you first feel it. Tell me when that point is reached, then I will increase it gradually beyond that to the point where you tell me to stop. You may tell me to stop when the shock begins to get unpleasant. It doesn't need to be painful; I don't want you to take all the shock you can. Each shock will last about a tenth of a second. Tell me when you first feel it."

After the shock was adjusted, the door to the shielded room was again closed and the subject was left alone.

III. After a two-minute wait the response to each tone was again tested twice in counter-balanced ABBA order.

IV. One minute after the last presentation of the tone, the first of 16 reinforced presentations of the two tones began. Each tone was reinforced with shock eight times, in the following order: BAABABBABAABABBA, counter-balanced.

V. Thirty seconds after the last reinforcement, a test of the conditioned response to the two tones was given in counter-balanced BAAB order.

VI. One minute after the last test the first of four additional reinforcements were given in counter-balanced ABBA order.

VII. Thirty seconds after the last reinforcement the extinction of the response to tone *A* (153 cycles) was begun. The tone was presented without reinforcement until the average magnitude of three successive responses was equal to or less than one third of the average magnitude of the responses to tone *A* in step V. The method of extinction to a criterion was used in order to obtain a measure of extinction comparable to the 'number of responses' measure in (4).

VIII. As soon as the criterion of extinction had been reached, the subject was taken out of the experimental room and given a 'rest interval' or recovery period of 5, 20, 60, or 180 minutes, according to the group into which he had been placed. Groups were balanced on the basis of two performance criteria: (i) mean amplitude of response to the two presentations of tone *A* in the test of conditioning in step V, and (ii) the number of presentations of the conditioned stimulus *A* in step VII (the first extinction) before the criterion of extinction was reached.

The subjects in the 5-minute interval group stood or walked around the room for three minutes of the interval and then returned to the chair in the shielded room where they spent the remaining two minutes. The subjects in the remaining groups spent the interval studying or reading in another room, and also returned to the experimental chair for the final two minutes of the period.

IX. The subjects were then given a second extinction series to measure recovery, in which tone *A* was presented without reinforcement until the mean amplitude of three successive responses was equal to or less than the mean amplitude of the two responses to tone *A* in the test of conditioning (step V). It should be noted that this criterion is the same as that used for the first extinction in step VII.

The interval between unreinforced presentations throughout the experiment averaged 30 seconds, and that between reinforced presentations averaged 60 seconds. The interval was varied when necessary to avoid presentation of the stimulus during or immediately following a spontaneous G.S.R.

The stimuli *A* and *B* were presented to half the subjects in each group in the order given above, and in the reverse order to the remaining half.

Eighty male subjects were used, 20 in each of the four groups. Five subjects in each group had previously taken part in a conditioned eyelid experiment. The others had not taken part in any experiments involving shock or conditioning. Very few knew anything of the nature of the G.S.R. before this was explained to them in step I. The explanation was given in order that the subjects would not expect shock from the galvanometer electrodes before the shock electrodes had been attached.

Results and Discussion

As already indicated, the four experimental groups were balanced on the basis of (i) mean *amplitude* of the first two conditioned responses to the 153 cycle tone following the 16 reinforcements and (ii) the *number* of unreinforced presentations of the tone required before the criterion of extinction was reached in the first extinction. The success of the balancing in terms of these two measures of the amount of conditioning may be seen from Tables 1 and 2. The differences between the group means are small and none is statistically reliable.

TABLE 1

Figures show for each group the means and σ_m's for the distributions of amplitude of the first two unreinforced responses to the 153 cycle tone following the reinforcement series (step V) and for the number of unreinforced presentations of the tone given before the amplitude of three successive responses in the first extinction decreased to one-third of the mean amplitude of the first two unreinforced responses. Groups are designated by the interval between the first and second extinction.

Measure	Group			
	5 minutes	20 minutes	60 minutes	180 minutes
Amplitude (millivolts)				
Mean	3.93	3.92	3.80	3.70
σ_m	.77	.75	.56	.56
No. responses				
Mean	23.60	21.50	21.50	22.85
σ_m	3.68	3.37	2.53	3.01

TABLE 2

Figures show the critical ratios of the differences between the means presented in Table 1

Group Means Compared	Critical Ratios for Amplitude	Critical Ratios for No. Responses
5 vs. 20 min.	.01	.42
5 vs. 60 min.	.14	.31
5 vs. 180 min.	.26	.16
20 vs. 60 min.	.13	.07
20 vs. 180 min.	.25	.30
60 vs. 180 min.	.14	.27

The recovery from the effects of the first extinction was measured by a second extinction, this extinction beginning

5, 20, 60, or 180 minutes after the criterion was reached in the first. Table 3 presents the mean amplitude of the first two responses in the second extinction for each subject, together with group means. Table 4 gives the number of presentations of the conditioned stimulus before the criterion was reached in the second extinction, and Table 5 the reliabilities of the differences between the group means presented in Tables 3 and 4. The critical ratios for the amplitude differences

TABLE 3

Figures show for each subject the mean amplitude in millivolts of the first two responses to the 153 cycle tone in the second extinction series (test of recovery). Groups are designated by the recovery interval used. Subjects are numbered consecutively beginning with 1 in each group.

Subject Number	Group			
	5 minutes	20 minutes	60 minutes	180 minutes
1	3.87	2.50	3.87	11.75
2	4.38	1.12	4.00	2.62
3	1.62	5.63	1.38	3.50
4	13.50	4.25	1.37	10.25
5	6.00	1.87	1.00	6.25
6	.62	5.13	4.88	.63
7	2.00	3.50	2.00	2.62
8	5.75	4.62	9.75	1.75
9	1.13	3.88	5.75	10.00
10	.50	3.50	2.25	.75
11	4.50	1.25	2.12	3.88
12	.75	4.12	13.50	13.00
13	2.00	1.88	3.13	4.37
14	3.50	7.25	9.75	.75
15	1.75	13.87	4.50	7.75
16	.62	2.75	11.00	2.00
17	.75	2.00	2.00	1.63
18	1.75	.88	4.25	9.87
19	1.38	12.62	8.62	5.00
20	5.50	3.50	.38	3.13
Mean	3.09	4.31	4.78	5.08
σ_m	.69	.78	.86	.89

were obtained in two ways: (i) without correlation, and (ii) with correlation used in computing the standard errors of the differences. The correlation of response magnitudes between groups for the second extinction was made possible by pairing subjects on the basis of a ranking in terms of the mean amplitude of the first two unreinforced responses to tone *A*

TABLE 4

Figures show for each subject the mean number of presentations of the conditioned stimulus before the criterion was reached in the second extinction series (test of recovery). Groups are designated by the recovery interval used. Subjects are numbered consecutively beginning with 1 in each group

Subject Number	Group			
	5 minutes	20 minutes	60 minutes	180 minutes
1	45	7	6	16
2	8	8	40	7
3	3	59	23	20
4	6	32	6	8
5	23	8	4	8
6	23	6	12	3
7	51	3	10	8
8	46	22	32	3
9	13	32	15	14
10	3	22	6	8
11	54	4	26	9
12	3	8	12	13
13	4	4	13	13
14	23	23	31	3
15	16	21	10	45
16	10	9	6	9
17	6	44	33	22
18	6	3	14	52
19	9	11	9	5
20	6	10	3	6
Mean	17.9	16.8	15.6	13.6
σ_m	3.87	3.40	2.50	2.93

TABLE 5

Figures show the critical ratios of the differences between the means presented in Tables 3 and 4. The standard errors of the differences for amplitude used in these computations were obtained in two ways: (i) without correlation, and (ii) with correlation of paired distributions. See text for the basis of the correlation

Groups Compared	Measure		
	Amplitude		Number of Responses
	(i) Uncorrelated	(ii) Correlated	
5 vs. 20 min.	1.17	1.37	.21
5 vs. 60 min.	1.54	2.60	.51
5 vs. 180 min.	1.76	1.78	.89
20 vs. 60 min.	.41	.54	.30
20 vs. 180 min.	.65	.78	.71
60 vs. 180 min.	.24	.23	.51

(step V in procedure). While none of the critical ratios are above 3.00, the smoothness of the curve (Fig. 1) suggests that the larger differences, at least, are probably real.

As the interval between the first and second extinction series was increased, the mean amplitude of the first two

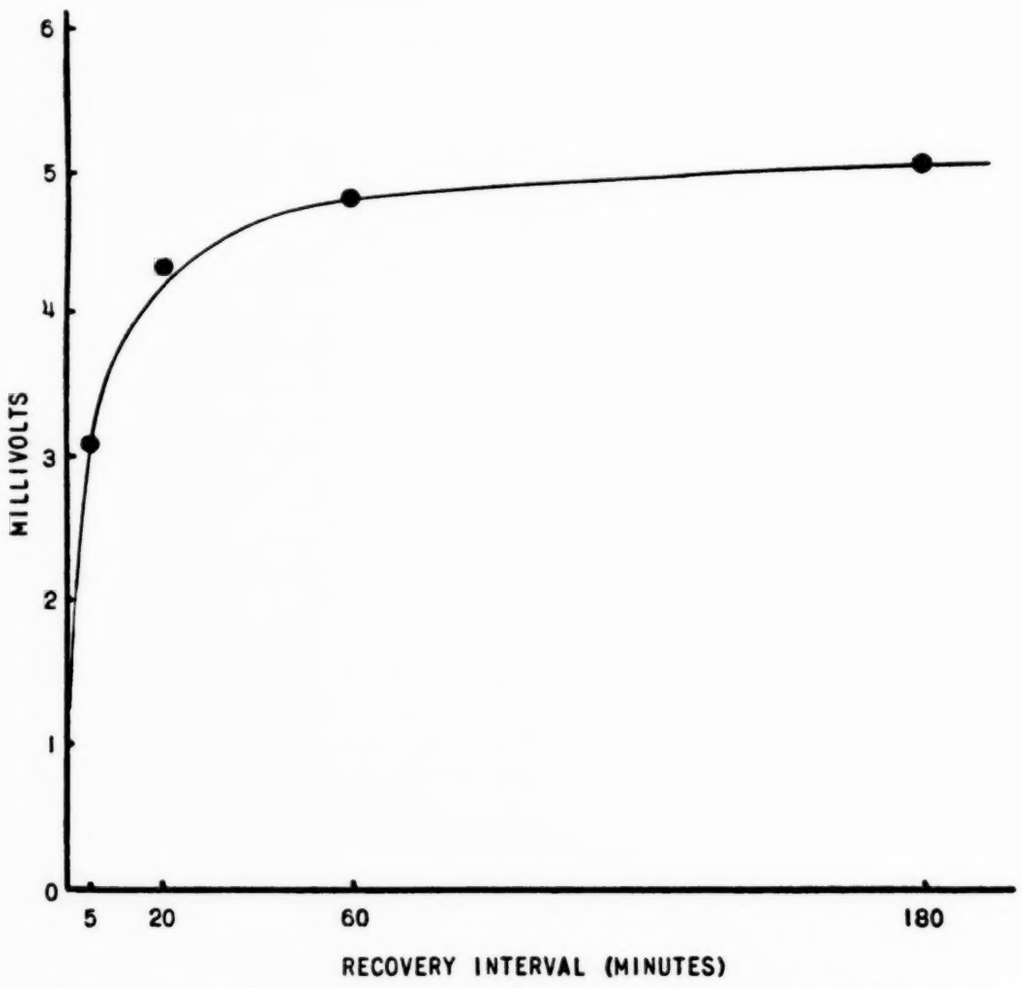


FIG. 1. Mean amplitude in millivolts of the first two responses in the second extinction as a function of the interval between the first and second extinctions. The smooth curve fitted to this data is plotted for the equation:

$$A = 5.25 - 4e^{-.1276 \cdot t},$$

in which A represents the amplitude of response, and t the length of the recovery interval.

responses in the second extinction series showed a progressive increase, with negative acceleration. The curve is very similar in general shape to the recovery curves for number-of-responses and time-for-extinction measures obtained with white rats in the preceding investigation (4, Experiment I).

The next fact to be noted is that, unlike amplitude, the number of presentations required for extinction to the criterion in the second extinction series shows no evidence of progressive increase as the interval between the two extinction series increases. In fact there is a slight decrease in number of responses with increasing interval, although none of the differences between means is statistically reliable. The means are presented in Table 4 and the critical ratios of differences in Table 5. The absence of any progressive recovery in terms of resistance to extinction is rather surprising, particularly since the amplitude measures obtained at the same time from the same subjects showed a definite curve of recovery. Furthermore, resistance to extinction, which does not show progressive recovery, is also the measure which is more comparable to the measures used in the previous animal experiment (**4**), which *did* show a progressive recovery. These results indicate that extreme caution must be used in generalizing the results of one experiment as a 'law of behavior,' either to another experimental situation or to a different organism.

The discrepancy between the results obtained with the two measures used in this experiment suggests that the conditioning of the G.S.R. is not a simple phenomenon in which a single factor which might be called 'response strength' increases or decreases with reinforcement, non-reinforcement, lapse of time, etc. At least two variables are involved, perhaps more.

The failure to find a progressive recovery using the resistance to extinction measure may be explained by the use of two assumptions: (i) resistance to extinction as measured under the conditions of this experiment is a measure of expectation of shock, and (ii) expectation (of shock) does not undergo a progressive change with the passage of time, at least within the limits involved in this experiment (5 minutes to 3 hours following extinction). The work of Cook and Harris (**3**) and Mowrer (**15**) makes the first assumption plausible. Mowrer found that under certain circumstances a subject will give a large response to a tone or light when he

has reason to expect a shock, and no response or a very minimal one when he does not expect shock. The gross changes in amplitude of response correlated with the two attitudes (sets), expectation or non-expectation of shock (**15**, p. 64) could very likely be sufficient to permit a subject to attain the criterion of extinction used in the present experiment (reduction of amplitude to one-third) when his set changed from expectation to non-expectation.[2] The second assumption is one for which there is no experimental evidence. The expectation judgment presumably depends upon verbal processes whose role in influencing behavior has received very little experimental investigation.

However, a technique recently described by Humphreys (**12**) offers promise of obtaining a direct check on the assumption that expectation does not change with time, although it sheds no light on the part played by speech. Using a simple apparatus consisting of two lights on a board, Humphreys asked subjects in a series of trials to predict whether the lighting of the left-hand light would be followed by the lighting of the right-hand light. By the use of this technique and obtaining a verbal or written report of the subject's expectation, it should be possible to determine whether there is a progressive recovery of an expectation which has been established by a series of paired presentations and then 'extinguished' by a series in which the second event did not occur.[3]

The hypothesis so far presented accounts for a failure to find an increase in resistance to extinction as the interval between the first and second extinction series is increased. However, the results of the experiment also showed a progressive increase with the length of the recovery interval, in the magnitude of the first two responses in the second extinction. To be consistent with the preceding assumptions, this change in amplitude of response as a function of the interval between

[2] Spontaneous remarks of a few subjects tended to support this view. Introspective reports were not systematically obtained, however.

[3] Since this was written an experiment has been performed with Humphrey's technique which confirms the assumption. No difference in expectation was found when 4 minute and 30 minute intervals were used (**13**).

extinctions cannot be explained as a change in expectation. Negative adaptation may be mentioned as an important factor affecting the amplitude of the G.S.R. It is suggested that the dissipation of negative adaptation effects during the recovery interval may account for the increase in amplitude of the initial responses of the second extinction series as a function of the lapse of time. Some justification for this latter assumption is found in data obtained by Hovland (**6**) in an experiment similar to the present one. Two groups of subjects were each given 24 tone-shock combinations in order to condition the G.S.R. The first extinction series was begun immediately after the last shock for one group of subjects and 30 minutes after the last shock for the other group. The mean amplitude of the first response was reliably greater for the second group. Since no previous extinction had occurred, this was presumably caused by recovery from negative adaptation (or 'inhibition of reinforcement').[4]

There is further evidence, from the present experiment, that negative adaptation is an important factor in influencing the amplitude of response for a given degree of expectation, *i.e.*, that expectation is not the only factor involved. For instance, expectancy alone could not account for the decline in amplitude of response during the series of tone-shock combinations. Humphreys (**12**), using the technique described above, found that if one event follows another consistently throughout a series of trials the expectancy of the second event progressively increases. This phenomenon is so common in everyday experience as to make experimental verification almost superfluous. Yet in the present experiment, where the shock followed the tone for 16 successive trials, the magnitude of the response showed a progressive *decrease*. The mean magnitudes of the G.S.R. following the 1st, 2d, 8th, 9th, 15th, and 16th paired presentations were 7·5, 7·4, 5·3, 5·2, 4·9, and 4·5 millivolts respectively. (The

[4] In the published article, the data were not presented in the form discussed here. Dr. Hovland kindly gave the writer access to the raw data from which the comparisons of the two groups were made. Unfortunately, it was not possible to obtain comparable measures of resistance to extinction.

remaining responses during paired presentation were not measured.) The conditions under which this steady decrement occurs suggest that the factor involved is negative adaptation. For a given level of shock-expectation, we may expect a considerable range of response amplitude depending upon the degree of negative adaptation.

The hypothesis, then, would be that the progressive recovery indicated by the amplitude measure represents the dissipation of negative adaptation effects accumulated during the training and first extinction. The recovery interval, being a break in the routine of extinction trials, resulted in a renewal of the expectation of shock, but the duration of this renewed expectation was not affected by the *length* of the interval. When the subjects again ceased to expect shock, a gross reduction in the response occurred, sufficient to permit the criterion of extinction to be reached.

One of the purposes of this experiment was to obtain a quantitative statement of spontaneous recovery as a function of the recovery interval, for use as a postulate in a deductive theory of behavior (**11**). The discrepancy between the results of this experiment and those of the previous animal experiment having the same aim (**4**) may be used as a criticism of this type of theory, on the grounds that in two situations, the same phenomenon may have entirely different quantitative characteristics. While this criticism may have some truth, it should be pointed out that the results of these two experiments are definitely not evidence for that point of view. Rather, they have shown that two phenomena customarily denoted by the same term may have entirely different characteristics. Operationally, then, they are different phenomena. The results of the present experiment, being apparently affected by more complex factors, may possibly be deduced from principles obtained from the simpler experiment (**4**) together with others. There is nothing in deductive theory to require that the behavior deduced by means of a given principle should resemble the behavior from which the principle was originally obtained.

It will be recalled that all of the subjects in the present

experiment were given eight presentations of the tone before it was paired with the shock, four presentations before the shock electrodes were attached, and four after. An attempt was made to assure the subjects that they need not expect shock during the time when first four tones were being presented. The nature of the G.S.R. was briefly explained to them in order that they would understand the necessity for the galvanometer electrodes and would not expect shock from that source. After the shock electrodes were applied, the shock was adjusted in intensity before the door to the shielded room was closed, and the subjects were told that they would receive further shocks. Then the second four tones were presented, unaccompanied by shock. It was supposed that by this procedure it would be possible to compare the response to the tones when the subjects did not expect shock with the response when they did expect it. However, the difference was not so marked as might have been expected from a consideration of the results of Cook and Harris (**3**) and Mowrer (**15**) mentioned above. The mean responses to the eight tones, measured in millivolts, before and after the electrodes were applied were as follows:

	Response Before Electrodes were Attached (Not Expecting Shock)				Response After Electrodes were Attached (Expecting Shock)			
Tone Number.....	1	2	3	4	5	6	7	8
Magnitude of response...	3.37	2.52	1.94	2.16	3.55	3.27	2.15	1.95

The differences in responses to tones in the corresponding positions in the two series are not statistically significant. The increase between the fourth and fifth responses is not trustworthy as a measure of response to increased expectation, since the four or five minute interval required for adjusting the shock between the fourth and fifth tones would permit recovery from any adaptation to the situation which had occurred, and also the shocks given during the determination of the desired intensity would presumably cause disinhibition of the adaptation (**16**).

The above evidence would seem to imply that expectation of shock does not significantly increase the magnitude of the galvanic skin response to the signalling stimulus. It certainly does not corroborate the findings of Cook and Harris (**3**) and Mowrer (**15**), that the G.S.R. can be conditioned verbally. However, observations of the reactions of individual subjects and conversations with a number of them have demonstrated the inadequacy of the technique. In the first place, the subjects were not sufficiently adapted to the situation before the tones were presented. Some of them expected shock from the recording (galvanometer) electrodes in spite of assurances that the electrodes were dead. Some subjects were so disturbed by the 'ice-box' appearance of the sound-shielded room that they gave extremely large initial responses to the tone, and others were upset by the appearance of the chair. One subject backed out of the experiment entirely as soon as he saw the chair, which he said reminded him of the 'hot seat.' For some of these extremely fearful subjects the administration of the shock during the adjustment of the intensity was actually reassuring. The shock itself was so mild in comparison to what they had apparently imagined might happen to them that they were considerably relieved by experiencing it. It was obvious that with many of the subjects the expectation of shock was not under the control of the experimenter. Considerably more rapport between subject and experimenter than was achieved at the time was necessary before assurances that shock would or would not follow could have any effect upon the expectation of the subject. From these considerations it is evident that the implications of any similarities or differences between the responses under the two conditions just described are inconclusive.

Summary

The galvanic skin response was conditioned to tonal stimuli and then extinguished in four groups of 20 subjects. After intervals of 5, 20, 60 or 180 minutes (a different interval for each group) spontaneous recovery was measured by means

of a second extinction. In terms of *mean amplitude of the first two responses* in the second extinction, a progressive, negatively accelerated recovery was found as a function of the increasing interval. Using the *number of responses required for extinction* to a criterion, however, no progressive recovery was found. An hypothesis is presented suggesting that these results may be explained in terms of two factors, (i) expectation of shock, which did not show a progressive alteration during the period following the first extinction, and (ii) negative adaptation, which influenced the amplitude of response obtained for a given degree of expectation.

An attempt to investigate the rôle of expectation of shock, planned as a part of the procedure, was found to be inadequate to the problem.

(Manuscript received June 6, 1939)

References

1. Burr, H. S., Lane, C. T., and Nims, L. F., A vacuum tube microvoltmeter for the measurement of bioelectric phenomena, *Yale J. Biol. and Med.*, 1936, **9**, 65–76.
2. Cole, L. E., A comparison of the factors of practice and knowledge of experimental procedure in conditioning the eyelid response of human subjects, *J. gen. Psychol.*, 1939, **20**, 349–373.
3. Cook, S. W., and Harris, R. E., The verbal conditioning of the galvanic skin reflex, *J. exp. Psychol.*, 1937, **21**, 202–210.
4. Ellson, D. G., Quantitative studies of the interaction of simple habits. I. Recovery from specific and generalized effects of extinction, *J. exp. Psychol.*, 1938, **23**, 339–358.
5. Hilgard, E. R., The relationship between the conditioned response and conventional learning experiments, *Psychol. Bull.*, 1937, **34**, 61–102.
6. Hovland, C. I., 'Inhibition of reinforcement' and phenomena of experimental extinction, *Proc. nat. Acad. Sci., Wash.*, 1936, **22**, 430–433.
7. Hovland, C. I., The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone, *J. gen. Psychol.*, 1937, **17**, 125–148.
8. Hovland, C. I., The generalization of conditioned responses: II. The sensory generalization of conditioned responses with varying intensities of tone, *J. genet. Psychol.*, 1937, **51**, 279–291.
9. Hovland, C. I., The generalization of conditioned responses: III. Extinction, spontaneous recovery, and disinhibition of conditioned and generalized responses, *J. exp. Psychol.*, 1937, **21**, 47–62.
10. Hovland, C. I., The generalization of conditioned responses: IV. The effects of varying amounts of reinforcement upon the degree of generalization of conditioned responses, *J. exp. Psychol.*, 1937, **21**, 261–276.

11. HULL, C. L., Mind, mechanism, and adaptive behavior, *Psychol. Rev.*, 1937, **44**, 1–32.
12. HUMPHREYS, L. G., Acquisition and extinction of verbal expectation in a situation analogous to conditioning, *J. exp. Psychol.*, 1939 (In press).
13. HUMPHREYS, L. G., MILLER, J., AND ELLSON, D. G., The effect of the intertrial interval on the acquisition, extinction, and recovery of a verbal expectation (In preparation).
14. JONES, H. E., Conditioned psycholgalvanic responses in infants, *Psychol. Bull.*, 1928, **25**, 183–184.
15. MOWRER, O. H., Preparatory set (expectancy) a determinant in motivation and learning, *Psychol. Rev.*, 1938, **45**, 62–91.
16. RAWDON-SMITH, A. F., Experimental deafness. Further data upon the phenomenon of so-called *auditory fatigue*, *Brit. J. Psychol.*, 1936, **26**, 233–244.
17. SKINNER, B. F., Two types of conditioned reflex and a pseudo-type, *J. gen. Psychol.*, 1935, **12**, 66–77.
18. SKINNER, B. F., *The behavior of organisms*. New York: D. Appleton-Century Co., 1938.
19. WILLIAMS, S. B., Resistance to extinction as a function of the number of reinforcements, *J. exp. Psychol.*, 1938, **23**, 506–522.
20. YOUTZ, R. E. P., Reinforcement, extinction, and spontaneous recovery in a non-Pavlovian reaction, *J. exp. Psychol.*, 1938, **22**, 305–318.

LEVEL OF ASPIRATION IN RESPONSE TO A PREARRANGED SEQUENCE OF SCORES[1]

BY JOHN W. GARDNER

Connecticut College

One of the earliest findings in the course of experimental work on level of aspiration was that marked individual differences in aspiration-level appear, even in the course of simple laboratory tasks. For example, one individual may exhibit a level of aspiration far above his level of past performance, even though his rate of improvement offers no grounds whatever for high expectations; another individual may habitually exhibit an aspiration-level so low that he is constantly provided with the presumably satisfying experience of doing 'better than he expected.' Hoppe (**9**), the pioneer investigator in the field, asserts that these individual differences in aspiration-level are diagnostic of important personality differences.

The question at once arises as to the extent to which these individual differences reflect a *general trait* of personality. Is the individual's response specific to the particular task involved, or would it exhibit consistency in a variety of unrelated tasks? Experimental evidence on the point has been conflicting. Frank (**3**) presents data which seem to indicate the presence of a general trait. Gould's study (**6**) fails to support these findings. The point is of obvious importance for the experimental study of personality. If it should prove that we are dealing with a general trait (a view supported by the results of the present investigation), the experimental psychologist will have an important addition

[1] The investigation here reported was carried out in the Psychological Laboratory at the University of California, and constitutes part of a dissertation submitted for the degree of Doctor of Philosophy. Many thanks are due Professor Warner Brown for keen and helpful criticism on matters of method and interpretation. Thanks are also due Professor R. C. Tryon for suggestions concerning statistical treatment of the data.

to his currently rather meagre supply of laboratory approaches to the study of personality.

The present study attempts (1) to devise an improved method of measuring aspiration-level behavior, (2) to examine the reliability of measures obtained with this method, and (3) to determine the generality of the behavior involved.

Method and Procedure

Frank, in his important first paper on the subject, has defined level of aspiration as "the level of future performance in a familiar task which an individual, knowing his level of past performance in that task, explicitly undertakes to reach" (3, p. 119). Chapman and Volkmann (2) have suggested that the term need not be limited to situations in which knowledge of past performance is involved. The present investigation follows Chapman and Volkmann in accepting Frank's definition with elimination of the past-performance clause. In this study the majority of statements concerning future performance are made after the subject has had experience in the activity involved, but the subject is also required to make certain statements concerning his aims *before entering the activity*, and the term level of aspiration is used here to cover both sorts of statement.[2]

The Prearranged Sequence of Scores.—In accordance with Frank's method, a score was reported to the subject after each trial in a given task, and he was then asked to state what score he intended to make on the next trial. But in the present experiment the scores reported to the subject were never his actual scores; after each trial the subject was given a score which was one of a *prearranged sequence of scores.* All of the subjects were provided with precisely the same prearranged 'performance curve' in a given task. The series of scores reported to all of the subjects in card-sorting is shown in Fig. 1. (The percentile scale used in connection with the scores will be explained later.) The important thing to note here is that this predetermined 'learning curve' was precisely the same for every subject in a given task. The *pattern* of the curve was also the same for all four tasks, but the pattern was carried out at different levels for the different tasks, so that the same actual scores would not be involved.

The obvious advantage of prearranged scores is the opportunity they offer of controlling the objective situation with which the subject is presented. There are reasons why such control of the objective situation is of considerable importance. Chiefly, it gives all subjects precisely the same objective basis for prediction, so that if one subject differs from another in the discrepancy between his level of aspiration and his level of past performance, this will not be attributable to differences in the objective situations, but can be traced unequivocally to the individual attitude or set which we infer to be back of aspiration-level behavior. A simple example will illustrate the difficulties which arise if this control is not achieved:

Suppose that instead of using a prearranged sequence of scores, we simply tell the subjects their true scores. One subject improves consistently, his scores running 10, 15, 20, 25. . . . Let us further suppose that this subject is highly realistic, and consistently bases his statements concerning future performance upon the objective

[2] The use of the term *level of aspiration* has been discussed at length in a previous paper (5).

evidence supplied him by his level of past performance. Under the circumstances, it is reasonable to suggest that his prediction for the fifth trial in the above series would approximate 30; and if the situation continued thus, his level of aspiration would average close to five points above his level of past performance, since the objective evidence available to him indicates that he is improving about that much on each succeeding trial. Now the most useful measure of aspiration-level behavior appears to be Frank's measure of the discrepancy between level of aspiration and level of past performance; and our subject's average score on this measure would be *plus five*. But now let us consider an equally objective individual whose scores run, let us say, 14, 15, 16, 15. . . . If this individual's aspiration-level behavior were governed by precisely the same sort of objective, realistic attempt to predict his performance as was operative in the first case cited, then his prediction on trial five would run very

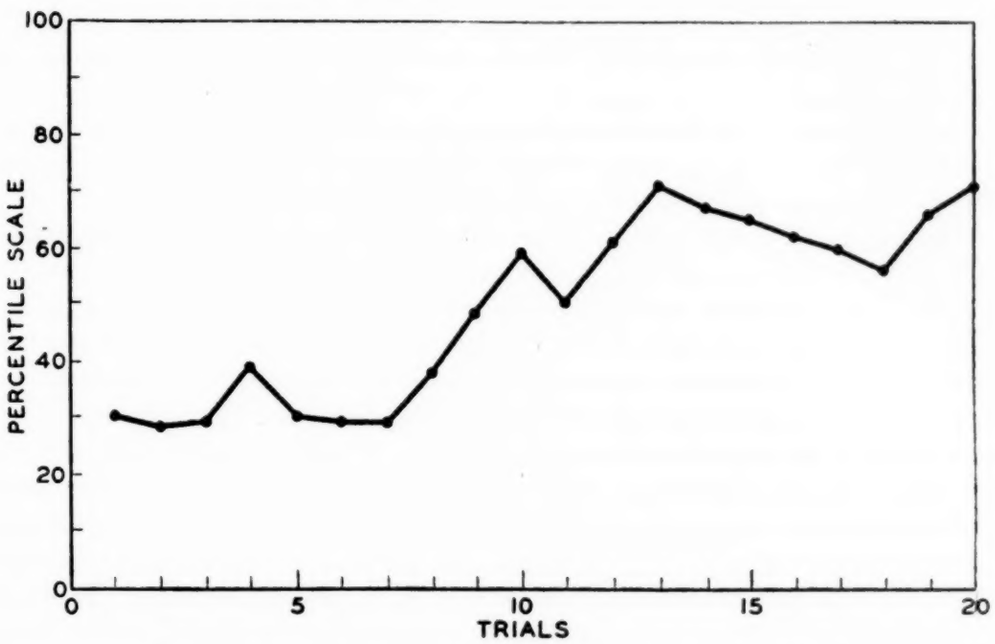


FIG. 1. The prearranged sequence of scores (card-sorting).

close to 15; and if the situation continued thus, his average discrepancy between aspiration-level and past performance-level would be close to zero. In short, we have here two individuals with precisely the same attitudes governing their aspiration-level behavior (*i.e.*, each assesses his potentialities for future performance on the basis of the only strictly objective evidence available—his level of past performance), yet these individuals behave in quite different fashions on our test, *because they are faced with quite different objective situations.*

However, quite aside from such suppositions, the present study will present empirical evidence that aspiration-level behavior does vary according to whether the performance curve is rising or falling. In short, both common sense and empirical evidence indicate that the pattern of the individual's performance curve up to the moment of prediction will influence his prediction; therefore, if we fail to make this pattern constant for all subjects, we cannot readily compare individuals in their reaction to that pattern.

After it had been decided that a prearranged performance curve common to all

subjects would provide the most feasible answer to the above problem, there remained the question as to what particular pattern this curve should follow. It was impossible to know what specific sequence of stimuli would prove most useful in evoking individual differences. However, it was thought that the following considerations could safely be advanced *a priori:*

1. The curve should remain on the same level for several trials after the task has begun. This seems desirable because the subjects are apt to come to the task with various hypotheses as to what their performance is *going to be*, and these hypotheses should be given a chance to iron out under the press of the objective situation before variations in the stimulus situation are introduced.

2. It would be desirable to make one section of the curve an unequivocally *rising* portion, so that individual differences evoked by such an encouraging state of affairs could be examined.

3. Another portion of the curve should be unequivocally *downward* in trend, so that the effect of this discouraging turn of events could be observed in different individuals.

4. The curve should have enough irregularity of pattern to simulate reality; and at the same time it must retain enough regularity for the subject to get an unmistakable impression of the trend of events.

Bearing in mind the few general considerations outlined above, and a few practical considerations such as length of time available, we worked out the pattern shown in Fig. 1. First, the subject was reported three scores at almost precisely the same level, allowing him time to obtain a clear-cut orientation to the level at which he was 'performing,' in accordance with the first consideration set forth above. The fourth score reported to the subject represented a striking improvement. It was thought that marked individual differences might be evoked by this single, sudden improvement; *i.e.*, some subjects might raise their aspiration-level sharply, others might await further evidence of improvement. Both Hoppe (9) and Jucknat (8) found evidence of such individual differences in response to a single improvement in score. On trials 5, 6, and 7, the performance curve drops back to its original level and stays there. It was hoped here that individual differences in aspiration-level behavior might appear when it becomes apparent that the improvement on trial 4 was only temporary. Some subjects might accept the situation realistically and adjust their aspiration-level downward. Others might refuse to lower their aspiration-level again. On trial 8 the curve begins to rise again, and continues to rise for three trials. It drops back for one trial, and then rises for two more, reaching a peak on trial 13. This portion of the curve represents a striking improvement in 'performance.' The single reversal in the middle of the series of improvements serves to break the pattern, in accordance with the fourth consideration set forth above; and there is the possibility that it will evoke significant individually different reactions in its own right. On trial 14 the curve starts downward, and continues thus for five successive trials. Coming at a time when the individual could reasonably expect to be reaching his peak performance, these slight but persistent drops in score were calculated to evoke individual differences in the response to a falling performance curve. Finally, two relatively good scores were reported to the subject to 'pick him up' and to send him away in a good humor. The response to these two trials did not enter into the data, but were designed for the sole purpose of ending the task on as neutral a note as possible. The final score was not precisely the same for all subjects but was varied within a range of seven points.

There was serious danger that the subjects would compare scores on the test, and it was found that the score most frequently remembered by the subject was the score on the final trial.

The question arises as to the extent to which the subject accepted the prearranged scores reported to him as actually his own scores. Unfortunately there does not seem to be any convenient way of studying this point systematically. It has been suggested that a study be made of the accuracy with which the subjects are able to estimate their own past performance. If they were quite accurate in doing so, the assumption would be that there is considerable risk of detection in reporting prearranged scores to them, the extent of the risk depending upon their degree of accuracy and upon the extent of the discrepancy between their actual scores and the scores reported to them. However, such an assumption would be unwarranted, since simple estimates of past performance would not be comparable to judgments in the present test situation. If the subjects want to arrive at an independent estimate of their past performance in the present experiment, they must do so in the face of a very strong positive suggestion (*i.e.*, the prearranged score reported to them by the experimenter). Thus the question is not simply how well the subject can judge his own performance, but how well he can judge his true performance in the face of a misleading suggestion. There seems to be no objective means of settling this question. The most that can be said is that on the basis of the experimenter's subjective impressions the prearranged scores seemed remarkably successful in gaining acceptance as actual performance scores.

The Percentile Scale.—The scoreboard pictured in Fig. 2 was used to indicate to the subject his score on a given trial, and at the same time was used by the subject in setting his aspiration-level. The scale is a percentile scale, and the setting of the right-hand marker by the experimenter conveyed to the subject not an absolute score but a statement of how well he did in relation to a group of his fellows. (Since all of the scores were prearranged, he was not, of course, actually being compared with a group of his fellows; the standing accorded him after a given trial was determined by the prearranged schedule of scores.) He was told that if the experimenter set the right-hand marker at 5 it meant that on that trial he had surpassed 5 percent of the group; 100 would mean that he had surpassed everyone; 50 would mean that he had just equalled the median score, *i.e.*, had surpassed 50 percent of the group. With the aid of such illustrations, the subjects apparently found no difficulty in grasping the percentile system of scoring.

As we have indicated, this scoring by the experimenter was carried on by means of the single marker on the right-hand side of the scoreboard. After the scoring system had been explained, the subject's attention was called to the two markers on the other side of the board. The upper marker was black, the lower one red. The subject was first instructed concerning the black:

> The group of students with whom you are being compared have all had twenty trials in this task. Before you start working, I want you to estimate how good you will be when *you* have had that much practice;—I want you to set the black marker to indicate the percent of the group which you expect to surpass when you have had twenty trials of practice. In other words, I want to know before you start how good you expect to be when you are at your best.

This estimate will be referred to as the ultimate level of aspiration. After setting this marker the subject is instructed further:

Now that you have indicated how high you expect to rise when you are at your best, I'd like you to set the *red* marker to indicate where you expect to stand on your *first trial* in the task. In other words, bearing in mind that the group with whom you are being compared have had 20 trials of practice, what percent of the group do you expect to surpass on *your* first trial?

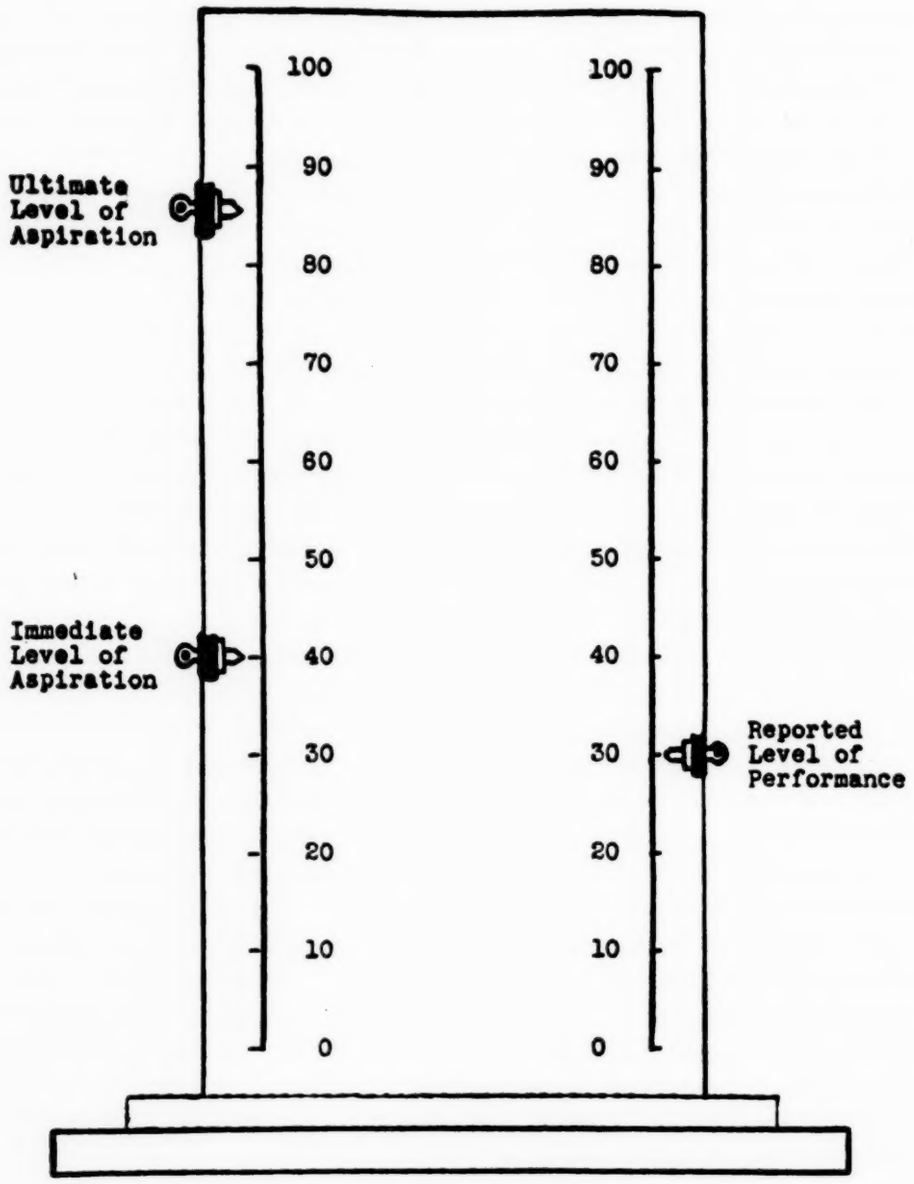


FIG. 2. Percentile scoreboard.

Note that (according to these instructions) the first-trial score of the subject is not being compared with the first-trial scores of his fellows but is being compared with the scores which his fellows have attained after twenty trials of practice. This basis of comparison remains the same throughout the test. The red marker is removed from the board after each trial, so that the subject is forced to make a fresh judgment each time. The estimate made with this red marker will be referred to as the immediate level of aspiration.

Previous investigators have not made use of such a percentile scale, and at first glance it may seem to be an unnecessarily complicated innovation. However, there is considerable reason to suppose that such a procedure is highly desirable. Hoppe, Frank and Jucknat all stressed the extent to which the subject evaluates his performance in social terms. For example, Jucknat points out:

> The subject does not by any means evaluate a task on the basis of its difficulty *per se*. . . . The subject experiences his performance not only as an objective triumph over difficulties but as a 'testing' in the eyes of the experimenter. He knows that his performance is compared with that of other subjects and that he receives a certain ranking in respect to them (**8**, p. 117).

In preliminary work leading up to the present investigation it became clear that subjects tend to judge themselves in terms of a rough sort of percentile scale. If no percentile scale is provided them, they make an effort to set up their own makeshift scale. They ask "How many get up this high?" or "Am I below average?" or "Does anyone ever get 'way up there?" Gould noted the occurrence of such questions as these, but felt that the social frame of reference could be abolished by refusing to answer such queries. However, if the experimenter refuses such information the subjects are simply left free to fall back upon hypotheses of their own, all of which (to the extent of the writer's observation and questioning) have to do not with the absolute goodness of their performance but with the relation of that performance to the group average or some similar reference point. In other words, the subject probably attempts to interpret any scores given him in terms of a social rank-order, whether we like it or not. If we ignore this we are simply leaving each subject free to form his own private hypotheses concerning the percentile value of a raw score; and there is good reason to suspect that such private hypotheses vary considerably from one subject to the next. This is a matter of importance, because it means that the subject is using a frame of reference over which we have not the slightest control, and which we actually know nothing about. The scores reported to the subject are the stimuli in our experiment, but the subject never reacts to the raw score *per se;* he reacts to 'what the score means to him.' And what it means to every subject is left to chance unless all subjects are provided with a standard frame of reference.

The Selection of Tasks.—The selection of appropriate tasks was a long and difficult process, and the final choice was a compromise between various competing necessities. The characteristics which make for desirability or undesirability in a task for such an experiment are numerous, ranging from the broadest sort of qualification to petty mechanical details. Some of the most important of these characteristics for our experimental situation were the following:

1. The task should be one in which the reporting of prearranged scores is feasible. *Time scores* lend themselves to such treatment, and consequently tasks which can be scored in terms of time suggest themselves as possibilities here. In some tasks it is difficult or altogether impossible to report prearranged scores. One cannot conveniently use prearranged scores on an individual who is shooting darts at a target.

2. It is desirable that the task be novel and relatively 'neutral' in the sense that it does not evoke *a priori* feeling of assurance or discouragement which may vary from subject to subject. Concretely, this means that the task should not be an obvious test of some familiar function in which the subjects are apt to regard themselves as having a special ability or disability (*e.g.*, rote memorizing, doing mathematical computations, spelling, and the like). Whatever the nature of the personality variables

which may be back of aspiration-level behavior, it is certain that familiarity with the task can have a very strong effect in limiting the estimates of future performance to within a certain range. Now if some subjects come to the task with incontrovertible evidence of their ability in that activity, and others come with incontrovertible evidence of their lack of ability in the performance, and still others just do not know where they stand, we are again faced with wide differences in the meaning of the situation to the individual—differences over which we have no control. If there is any generalized trait back of aspiration-level behavior it can fairly be expected to appear only if all subjects enter the test with approximately the same amount of specific experience in the task involved, and approximately the same *kind* of experience in that activity. There is only one way to achieve this difficult goal, and that is to attempt to set up tasks in which none of the subjects has had any specific experience with the activity involved.

3. The task should be such that quite brief trials are feasible (because of our particular time limitations); and it should be such that it may be divided up into a considerable number of trials.

4. Tasks should be such that the performance does not vary greatly from trial to trial. This is essential to the reporting of prearranged scores. The reference here is to such extreme variations as occur from time to time in the block-design test (which was tried out and rejected), in which the subject will experience some sort of mental 'blocking' or will 'blow up' completely on a trial now and then. Similarly in anagrams (which was also tried and rejected) the subject will occasionally give up completely on a particular trial. Under such circumstances the subject knows that he has failed miserably on that trial and no prearranged score could possibly convince him otherwise.

5. The tasks used should, as far as possible, sample different abilities.

These were the major considerations. After a preliminary survey of possible tasks, twenty were selected for an actual try-out. It would not be practicable to recite here in detail all of the difficulties met in trying out these twenty tasks. Four tasks were finally selected: card-sorting (7), digit-symbol substitution, a multiple-choice test of opposites, and cancellation performed while counting backward by 3's. These tasks suited the requirements outlined above in most respects. The test of opposites seems to conflict with the requirement that the tests used should not be obvious tests of some familiar function in which the subjects would be apt to regard themselves as having a special ability or disability, but the test was so satisfactory in other respects that it was retained.

Subjects.—The subjects used were 31 male college students taken from the elementary psychology course, sophomores and juniors for the most part. Participation in psychological experiments was a uniform requirement for all students in the course, so a relatively unselected sampling of this group was obtained. To what extent the elementary psychology enrollment is representative of the college population in general remains an open question. In any case it is probably particularly true in this realm of investigation that the 'average college student' is not the 'average human' even though the level of actual ability does not enter into this study. The college student is pretty well oriented to competition in almost any sort of test situation. He is accustomed to viewing his performance in its relation to a group average. And this relationship between his performance and the group average is of relatively high importance to him.

Each subject was tested alone. In the first session the subject performed the

card-sorting, and then immediately afterwards he performed the cancellation and counting backwards. A week to two weeks later he was called in for a second session and given the digit-symbol substitution and the opposites test.

Data and Results

It was apparent from the first that the technique was evoking wide individual differences in response. Figure 3 is a graphic record of an individual with an unusually high level of aspiration. Even before this subject has begun the task, he predicts that he will surpass 40 percent of the group on his first trial. When he is told that he surpassed 30 percent

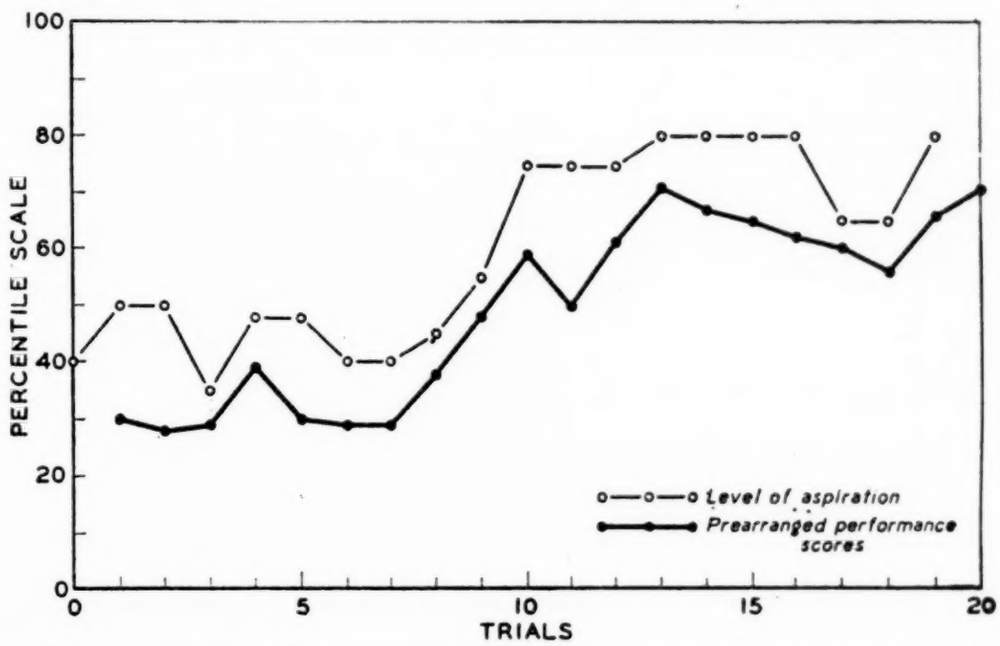


Fig. 3. Subject with consistently high level of aspiration (card-sorting). Each filled dot represents a score reported to the subject. The unfilled dot *on the same vertical ordinate* represents the subject's statement of what he intends to do on the subsequent trial. The discrepancy between the filled and unfilled dots on the same vertical ordinate is the basis for the aspiration-level score.

of the group on that trial, he promptly raises his estimate to 50 percent for the second trial, and leaves it there even though this second trial turns out to be less than 30. When the third trial also proves to be less than 30, he drops his aspiration-level for the fourth trial to 35, still six points above his performance level. Throughout the performance the subject exhibits a large positive discrepancy between the score attained on a given trial and the aspiration-level for the succeeding trial.

Figure 4 presents the record of an individual whose aspiration-level is rather consistently low. In the early trials when the performance is on a very low level, he ventures a few estimates which are above his past-performance level; but thereafter his estimates fall regularly below the performance curve.

Scoring Method.—There was no way of knowing *a priori* what would be the most fruitful way of scoring such records as these. There was not even any way of knowing *a priori* whether the behavior of the aspiration-level at one part of

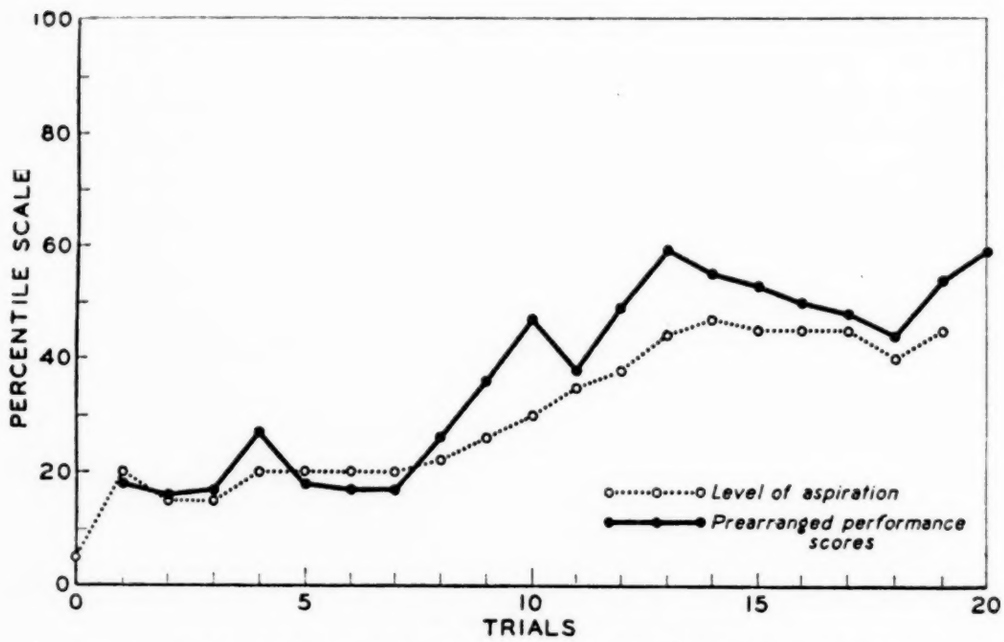


FIG. 4. Subject with consistently low level of aspiration (opposites test).

the curve correlated positively with aspiration-level behavior at any other part of the curve. Therefore, a number of different scores were tried out as an exploratory measure.

Figure 5 presents the aspiration-level record of a hypothetical individual and indicates the different scores used. *Score 1* was the subject's aspiration-level for the first trial.[3] This initial level of aspiration was set before he had any experience in the performance. *Score 2* consisted of the discrepancy between the performance on trial 3 and the aspiration-level for the succeeding trial (note that the performance

[3] Thus for the hypothetical individual in Fig. 5, score 1 would be 50.

on a given trial and the aspiration-level for the *subsequent trial* are plotted on the same vertical ordinate in Fig. 5).[4] It was thought that score 2 might provide a measure of how

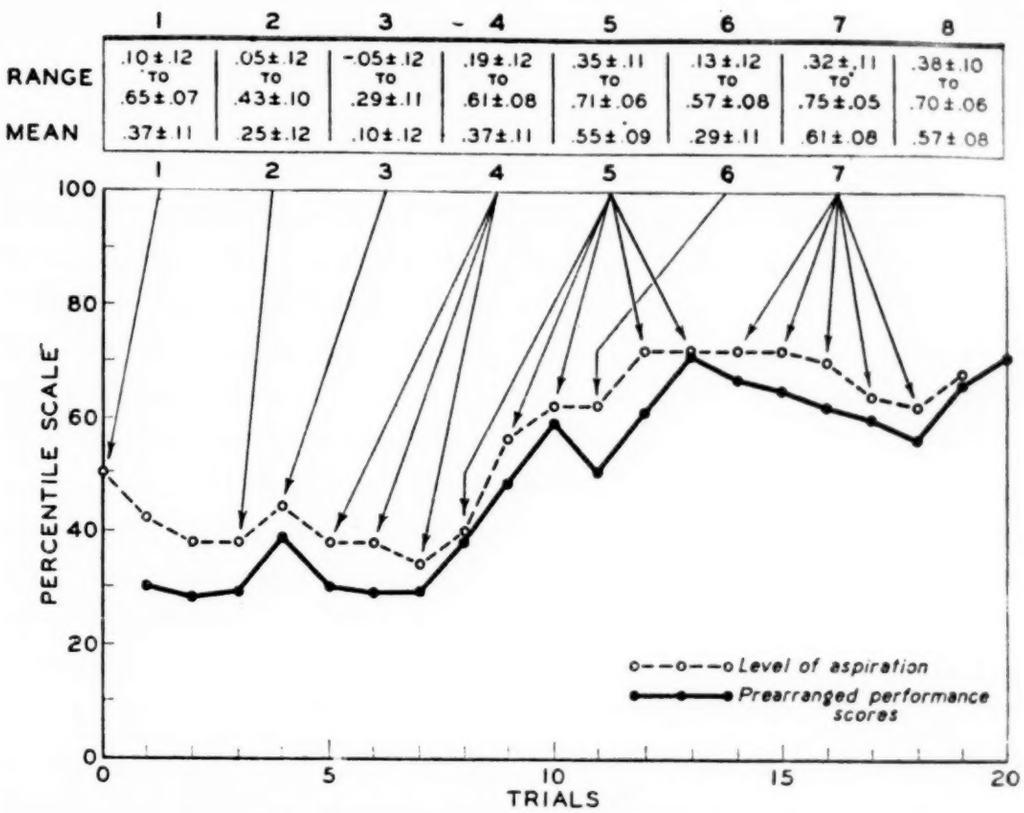


FIG. 5. The scores used, with intercorrelations between tasks for these scores. *Score 1* is the S's aspiration-level for the initial trial. This aspiration-level is set before the first trial and is plotted on the zero ordinate. *Score 2* is the discrepancy between the score on trial 3 and the aspiration-level for trial 4 (plotted on the same vertical ordinate). *Score 3* consists of the difference in aspiration-level immediately preceding and immediately following trial 4. *Score 4* is the average discrepancy between aspiration-level and past performance level over trials 5, 6, and 7. *Score 5* is the same average discrepancy over trials 8, 9, 10, 12, and 13. *Score 6* is the difference in height of aspiration-level immediately before and immediately after trial 11. *Score 7* is the average discrepancy between aspiration-level and past performance level over trials 14, 15, 16, 17, and 18. *Score 8* is the same average discrepancy over trials 3–18 (*i.e.*, the entire curve with the exception of the first two and last two trials).

well the subject had adjusted his aspiration-level to the level at which he was performing, after the first three trials had provided him with unequivocal evidence on this latter point.

[4] For the hypothetical individual in Fig. 5, score 2 would consist of the discrepancy between 29 (performance on trial three) and 38 (aspiration-level for subsequent trial), viz., nine points.

Score 3 attempts to measure the effect upon the aspiration-level of the striking improvement on trial four. Several methods of measuring this were tried out, and the measure finally decided upon was the difference between the aspiration-level immediately before the improvement occurred and the aspiration-level immediately after it had occurred.

After this sudden improvement, the curve drops back to its former level, and stays there for three trials. How will the subject alter his aspiration-level when it becomes apparent that the improvement on trial four was only temporary? Some subjects promptly lower their aspiration-level. Others refuse to lower it. Still others adjust it gradually downward over the course of three trials. *Score 4* is an attempt to measure the extent to which the subject does or does not adjust his aspiration-level downward on this part of the curve, and consists of the *average discrepancy* over these three trials between aspiration-level and level of past performance. *Score 5* consists of the same sort of average discrepancy on the five improvement trials in the middle portion of the curve, and is an attempt to measure the individual's alterations in aspiration-level when his score is improving rapidly.

But it will be noted that in the series of improving scores there is a single reversal. The individual's reaction to this reversal is measured in *score 6*. Several methods of scoring this reaction were tried, and the measure finally used was simply the difference between the height of aspiration-level before and after the reversal occurred.

The same sort of average discrepancy score which was used in scores 4 and 5 was used for the series of repeated score decrements near the end of the performance (trials 14, 15, 16, 17, and 18). The discrepancies between aspiration-level and level of past performance for these five trials were averaged to obtain *score 7*.

Score 8 is simply the average discrepancy between aspiration-level and past performance level for the entire curve with the exception of the first two and last two trials. The initial level of aspiration does not enter into this score, since at the time it was obtained the subject had no level of past per-

formance to estimate from (therefore, a discrepancy score could not be calculated). The first two trials and the levels of aspiration based upon them were not included in score 8 because it was believed that the subject had not yet had sufficient opportunity to orient himself to the situation. And, as has been said before, the last two trials served as 'consolation' trials and nothing more. Therefore, score 8 involves trials 3–18 inclusive.

Intercorrelations between Tests.—For each of these measures, all possible intercorrelations between tasks were worked out to get an estimate of the degree of generality of the behavior involved in these scores.[5] Table I presents these intercorrelations.[6]

Low intercorrelations between tasks appear on all of the scores which were based upon single measurements (scores 1, 2, 3 and 6). Whether or not these low intercorrelations are traceable to the unreliability of these single measurements cannot be determined, since reliability coefficients cannot be computed on the scores involved.

Of the scores based upon more than one measurement, score 4, consisting of three measurements, was the poorest, with a range of intercorrelations from .19 ± .12 to .61 ± .08, and an average intercorrelation of .37. Scores 5 and 7, each based upon five measurements, showed remarkably similar intercorrelations. Score 5 showed a range from .35 to .71 with an average of .55. Score 7 showed a range from .32 to .75 with an average of .61. Score 8, based upon sixteen measurements, showed intercorrelations closely resembling those of scores 5 and 7. The range was from .38 to .70, with an average at .57. The intercorrelations between tasks on scores 5, 7, and 8 indicate a considerably higher degree of generality than is usually found in personality variables.

It will be remembered that before the subject began the task he was asked not only to estimate his performance on

[5] All of the correlations were obtained by the product-moment method.

[6] Since it is not easy for the reader to keep in mind what these scores consist of, it may help to refer to Fig. 5; above each score in Fig. 5 is the range of intercorrelations between tasks for that score, and the average of these intercorrelations.

the first trial but also to estimate the highest performance level he would reach in twenty trials. This latter was termed the ultimate level of aspiration. As in the case of scores 1, 2, 3, and 6, the score on ultimate level of aspiration was based upon a single measure; so there was reason to doubt that the intercorrelations between tasks would run very high. How-

TABLE I

INTERCORRELATIONS BETWEEN TASKS ON EACH OF THE EIGHT SCORES

		Cancellation	Substitution	Opposites
		r P.E.	r P.E.	r P.E.
Score 1	Cards	.58 ± .08	.36 ± .11	.10 ± .12
	Cancell.		.27 ± .11	.26 ± .11
	Subst.			.65 ± .07
Score 2	Cards	.37 ± .10	.43 ± .10	.05 ± .12
	Cancell.		.11 ± .12	.13 ± .12
	Subst.			.38 ± .10
Score 3	Cards	.12 ± .12	.10 ± .12	.03 ± .12
	Cancell.		.05 ± .12	.29 ± .11
	Subst.			.12 ± .12
Score 4	Cards	.31 ± .11	.42 ± .10	.30 ± .11
	Cancell.		.19 ± .12	.39 ± .10
	Subst.			.61 ± .08
Score 5	Cards	.71 ± .06	.44 ± .10	.35 ± .11
	Cancell.		.57 ± .08	.65 ± .07
	Subst.			.57 ± .08
Score 6	Cards	.21 ± .14	.20 ± .12	.13 ± .12
	Cancell.		.57 ± .08	.17 ± .12
	Subst.			.46 ± .10
Score 7	Cards	.70 ± .06	.46 ± .10	.32 ± .11
	Cancell.		.71 ± .06	.75 ± .05
	Subst.			.69 ± .06
Score 8	Cards	.69 ± .06	.41 ± .10	.38 ± .10
	Cancell.		.54 ± .09	.70 ± .06
	Subst.			.67 ± .07

ever, as can be seen from Table II, these intercorrelations are remarkably high, comparing favorably with the intercorrelations on scores 5, 7, and 8.

Reliability of Scores.—It was possible to compute reliability coefficients on scores 4, 5, 7, and 8. On scores 4, 5, and 7, there was some doubt as to how valid an estimate of

reliability could be obtained with an odd-even or split-half technique, so an alternative technique was employed. All of the measures which went to make up the score were intercorrelated, and then the average of these intercorrelations was corrected by the Spearman-Brown formula to provide our final estimate of the score's reliability. Figure 6 presents the reliabilities thus obtained. The reliabilities on score 8 were obtained by a modified odd-even technique. The sixteen measures which made up score 8 were divided into the conventional odd and even sets; but the odd contained only

TABLE II

INTERCORRELATIONS BETWEEN TASKS ON ULTIMATE LEVEL OF ASPIRATION

	Cancellation	Substitution	Opposites
	r P.E.	r P.E.	r P.E.
Card-sorting	.44 ± .10	.58 ± .08	.35 ± .11
Cancellation		.51 ± .09	.50 ± .09
Substitution			.75 ± .05

two trials in which the subject's score was improving, whereas the even set contained four such trials. To balance the two sets in this respect, the first odd trial (a trial showing no improvement) was thrown in with the even set, and the first even trial was thrown in with the odd set. Thus the makeup of the two sets to be correlated was as follows:

A: trials 4, 5, 7, 9, 11, 13, 15.
B: trials 3, 6, 8, 10, 12, 14, 16.

The coefficients of correlation between these two sets of trials on all of the four tasks are presented in Fig. 6. The Spearman-Brown formula has been applied to the raw correlations to obtain the coefficients in the figure.

The computation of reliabilities on scores 4, 5, 7, and 8 permitted the correction for attenuation of the intercorrelations between tasks involving these scores. Figure 6 presents the range and average of these corrected coefficients.

Correlation between Different Portions of the Curve.—Table III presents the coefficients of correlation between scores 4,

5, and 7. It will be remembered that one of the reasons that it was necessary to assign separate scores to different portions of the curve was that there was no other way of knowing to

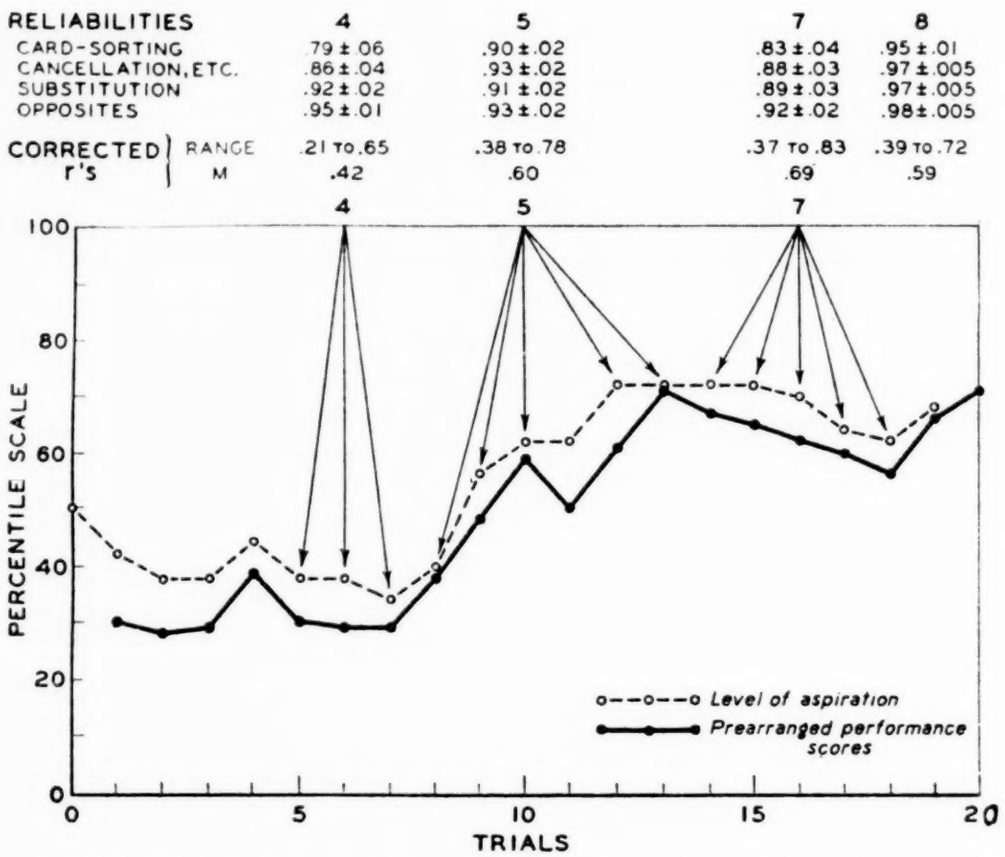


FIG. 6. Reliability coefficients on scores 4, 5, 7, and 8, with intercorrelations between tasks corrected for attenuation.

TABLE III

INTERCORRELATIONS BETWEEN SCORES 4, 5, AND 7

	4 *vs.* 5			4 *vs.* 7			5 *vs.* 7		
	r	P.E.	Corr'n for Atten.	*r*	P.E.	Corr'n for Atten.	*r*	P.E.	Corr'n for Atten.
Cards	.39	± .11	.47	.53	± .09	.65	.62	± .08	.72
Cancellation	.60	± .08	.67	.76	± .06	.87	.61	± .08	.68
Substitution	.43	± .10	.47	.57	± .08	.63	.70	± .06	.78
Opposites	.56	± .09	.60	.75	± .05	.80	.80	± .04	.90

what extent these different portions of the curve were correlated. There was no way of knowing to what extent aspiration-level behavior during improvement in performance was

correlated with aspiration-level behavior during decrement in performance-level. With scores on the different portions of the curve, we were in a position to compute these correlations. The portions of the curve dealt with were those covered by scores 4, 5, and 7. The correlations indicate a tendency for individuals to maintain the same position relative to the group under the varying conditions of performance.

Size of Discrepancy Scores on Different Portions of Curve.—Although Table III indicates that individuals tend to maintain approximately the *same rank within the group* under the varying conditions of performance, the group as a whole shows a strong tendency to respond differently to the different portions of the curve. This tendency is clearly indicated in Table IV, which presents the average discrepancy scores on the

TABLE IV

MEANS AND S.D.'S OF THE DISTRIBUTIONS OF SCORES 4, 5, 7, AND 8*

	4		5		7		8	
	M	S.D.	*M*	S.D.	*M*	S.D.	*M*	S.D.
Cards	8.50	4.16	0.60	3.10	6.18	3.48	4.70	3.32
Cancellation	5.18	3.10	0.69	5.10	4.95	3.84	2.76	3.32
Substitution	9.27	4.35	1.62	4.39	6.60	4.00	5.08	3.52
Opposites	8.76	4.71	0.47	4.56	5.85	4.45	4.57	3.81
Average on all tasks	7.30	4.97	0.50	3.72	5.60	3.21	4.34	2.90

* Scores 4, 5, 7, and 8 all measure discrepancy between level of aspiration and level of past performance. Scores 4, 5, and 7 cover different portions of the curve, and score 8 measures the average discrepancy for the curve as a whole.

various portions of the curve and for the curve as a whole. The highest discrepancies between aspiration-level and level of past performance occur in score 4, *i.e.*, at an early period in the performance when performance scores are still on a low level. Then as the subject begins to improve rapidly, the discrepancies drop almost to zero. Note that the aspiration-level itself is not falling during this period; the aspiration-level is rising, but it is not rising as rapidly as the performance curve. Finally, during the period of repeated decrements in performance score the discrepancies again increase markedly.

This tendency for the discrepancy between aspiration-level and past performance level to be markedly greater when the performance curve is falling than when it is rising is pictured in Fig. 7.

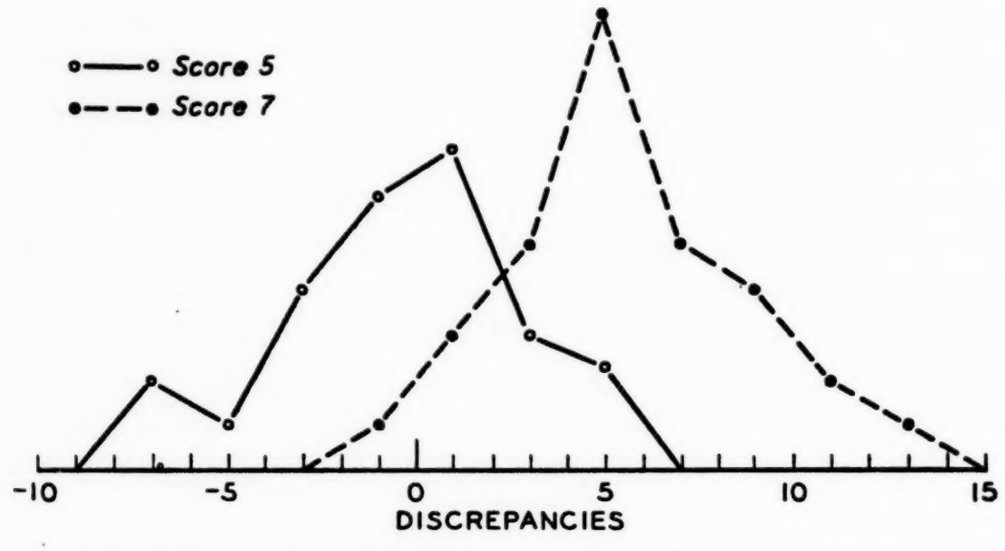


FIG. 7. Distributions of scores 5 and 7. Score 5 is a measure of the discrepancy between aspiration-level and level of past performance when the performance scores are improving rapidly. Score 7 measures the same discrepancy when the performance scores are falling off.

CONCLUSIONS AND DISCUSSION

1. The methodological innovations have proven themselves workable. The purely technical and mechanical difficulties involved in presenting a series of prearranged scores can be readily overcome; and the subjects showed themselves quite willing to accept a percentile scale for scoring and setting aspiration-level.

2. Such an experimental situation and technique does evoke marked individual differences in aspiration-level behavior, and these differences may be reliably measured.

3. The behavior underlying these measures is not specific to the task in question, but is general to the extent that rather high intercorrelations appear between tasks which are quite different in nature. This finding supports the results which Frank obtained, and conflicts with the findings of Gould. The measures in the present study which are most comparable to those obtained by Gould are the measures rep-

resented by score 8. The intercorrelations between tasks on score 8 averaged .57 ± .08. Gould's intercorrelations between tasks had a median of .29. Gould concluded on the basis of her experiment that generality could not be demonstrated with functionally unrelated materials. Our own results do not demonstrate complete generality; and there is at least a possibility that such generality as was obtained is only generality within a rather broad sphere of 'psychological testing'; *i.e.*, it may be that the generality here obtained is due to a tendency on the part of the subjects to assign a common significance to all psychological tests, or to all tests-taken-in-the-psychology-laboratory, of whatever nature. But results of the present study at least indicate generality within this rather broad sphere; Gould found very little generality working in precisely the same sphere. It seems reasonable to assume that these divergent results may be traced to differences in approach to the problem and differences in experimental technique. Among the most important of these differences are the selection of tasks and the use in the present investigation of prearranged scores and a percentile scale of scoring. To what extent these differences actually do account for the divergent results is of course a matter of conjecture. Our own beliefs concerning the methodological requirements for solution of the problem of generality have been stated in the early part of the present paper.

4. Discrepancy scores on different portions of the performance curve show rather high positive correlations.[7] This means that the standing of the subjects in relation to one another remains much the same under the varying conditions of performance. This finding may easily lead to the false supposition that prearranged performance scores are unnecessary, since whether the performance curve is rising or falling does not affect the individual's standing in the group. Whether the performance curve is rising or falling does not affect the individual's standing in the group *provided* that all

[7] This finding is corroborated by similar evidence from another study by the present writer, in which the same technique was used with adolescent subjects in an attempt to examine the relation of certain personality variables to level of aspiration (4).

the rest of the group is undergoing the same experience at the same time. Rejecting prearranged scores would mean that one individual's performance curve might be rising at the same time that another's might be falling.

5. The size of the discrepancy scores varies on different portions of the performance curve. The discrepancy scores decreased during the period when the performance was improving rapidly and increased during the period when the performance was falling off. This finding was incidental to the purpose of the present investigation; and this experiment was not so designed as to permit any generalizations on the matter. The point is reported because it appears to supplement (and corroborate) certain findings already reported by Sears (**10**) and by Anderson and Brandt (**1**) concerning the effect of 'success' and 'failure' upon level of aspiration.

Summary

Thirty-one college men were tested on each of four different tasks. Each task consisted of twenty trials. In accordance with Frank's method of testing level of aspiration, a score was reported to the subject after each trial in a given task, and he was then required to state a level of aspiration for the succeeding trial. In this experiment, the scores reported to the subject were never his actual scores; every subject in the group was presented with the same series of prearranged scores. The *pattern* of this prearranged performance curve was also the same for all four tasks, although the curve was shifted upward or downward a few points in the different tasks so that the same actual numbers would not be involved. The prearranged scores were reported in terms of a percentile scale, and the subjects stated their aspiration-levels in terms of the same percentile scale. The four tasks used were card-sorting, digit-symbol substitution, and multiple choice test of opposites, and cancellation performed while counting backward by threes.

The chief aspect of aspiration-level measured was the discrepancy between aspiration-level and level of past performance. It was found that highly reliable measures could be

obtained, and high intercorrelations between tasks indicated that the behavior underlying these measures was characterized by a considerable degree of generality.

Discrepancies between aspiration-level and past performance level tended to be positive and low when performance was improving rapidly and markedly higher when the performance curve was falling. Individuals tended to maintain the same rank within the group with respect to discrepancy scores when the performance curve was rising as when it was falling.

(Manuscript received September 21, 1939)

References

1. Anderson, H. H., and Brandt, H. F., Study of motivation involving self-announced goals of fifth grade children and the concept of level of aspiration, *J. soc. Psychol.*, 1939, **10**, 209–232.
2. Chapman, D. W., and Volkmann, J., A social determinant of the level of aspiration, *J. abnorm. soc. Psychol.*, 1939, **34**, 225–238.
3. Frank, J. D., Individual differences in certain aspects of the level of aspiration, *Amer. J. Psychol.*, 1935, **47**, 119–129.
4. Gardner, J. W., The relation of certain personality variables to level of aspiration, *Psychol. Bull.*, 1939, **36**, 540.
5. Gardner, J. W., The use of the term level of aspiration, *Psychol. Rev.*, 1939, **46** (in press).
6. Gould, R., An experimental analysis of 'level of aspiration,' *Genet. Psychol Monogr.*, 1939, **21**, 1–116.
7. Husband, R., and Miles, W., On sorting packs of 60 cards with form and color as variables in 2 to 6 kinds; card-sorting by reaction to the previous card, *J. appl. Psychol.*, 1927, 465.
8. Jucknat, M., Leistung, Anspruchsniveau und Selbstbewusstsein, *Psychol. Forsch.*, 1937, **22**, 89–179.
9. Hoppe, F., Erfolg und Misserfolg (Untersuchungen zur Handlungs und Affektpsychologie. Ed. by Kurt Lewin), *Psychol. Forsch.*, 1930, **14**, 1–62.
10. Sears, P. S., Levels of Aspiration in Academically Successful and Unsuccessful Children. Ph.D. Dissertation, 1939. (On file at Yale University Library.)

EXPERIMENTAL STUDIES IN ROTE-LEARNING THEORY. V. COMPARISON OF DISTRIBUTION OF PRACTICE IN SERIAL AND PAIRED-ASSOCIATE LEARNING

BY CARL IVER HOVLAND

Department of Psychology, Institute of Human Relations, Yale University

Introduction

A decrease in the rate of presentation of syllables in serial learning has been shown to be accompanied by a decrease in the amount of reminiscence obtained (**5**). It has also been found that material learned at a rate which shows no reminiscence requires as many trials to learn by distributed as by massed practice, but material learned at a rate where reminiscence is obtained is greatly benefited by distributed practice (**6**). The problem arises as to whether this relationship between reminiscence and distribution of practice efficacy is a general one—whether distributed practice is only efficient with conditions and materials which produce reminiscence. This is the contention of Bunch and Magdsick (**1**). A comparison between the efficacy of distributed practice with serial and with paired-associate materials would give evidence on this point, since a recent study by the writer (**7**) showed that paired-associate materials give much less reminiscence than serial lists learned under identical conditions. It would be predicted from the above relationships that paired-associate learning, which shows little reminiscence, would be much less benefited by distribution of practice than serial learning, where reminiscence is more pronounced. This prediction was tested in the present experiment.

Experimental Procedure

Subjects learned on successive days both paired-associate and serial lists by massed and by distributed practice. The relative efficacy of distribution of practice for the two types of learning could thus be determined. To permit satisfactory comparison,

the difficulty of the two types of materials was equated in a preliminary experiment (*cf.* 7). The results indicated that approximately the same number of trials are required to learn eleven units by the serial method as are required to learn nine pairs by the paired-associate method. These two lengths were consequently employed in the present experiment.

The details of the experimental procedure have been previously described (4, 7). The Hull exposure apparatus was employed. The experiments were performed in a sound-proof chamber to eliminate distractions. The anticipation method was used for both types of learning. The customary rules for construction of nonsense syllable lists were followed. For the paired-associate procedure the bands were prepared in such a way that the succession of the pairs was changed on successive presentations. Four experimental and two practice bands were employed for each learning method.

During massed practice only six seconds elapsed between successive presentations of the list. During distributed practice, a learning trial and a rest pause were regularly alternated. The rest pauses were two minutes in length. During the pause the subjects named colors automatically presented on the memory drum.

The rate of presentation was two seconds per syllable for both learning methods. This meant that in the paired-associate method four seconds were devoted to each pair. The instructions used with serial learning were identical to those given in (4); those used with paired-associates were the same as those given in (7). The schedule was prepared in advance in such a way that balanced practice orders were employed throughout to equalize practice effects and average out constant errors.

Thirty-two Yale college students, paid for their services, were employed. They had all been employed previously in memory experimentation and consequently only required four practice periods (one for each experimental program). During the main experiment the subjects learned two lists with each procedure, a total of eight lists in all. The subjects reported at the same hour every day except Sunday. They did not know the purpose of the experiment nor were they told in advance which experimental program would be employed on a given day.

Results

Data on the number of trials required to learn to the criterion of one perfect recitation by massed and distributed practice with paired-associate and with serial learning methods are given in Table 1. It will be observed that distribution of practice shows its customary effectiveness in reducing the number of trials required for learning in the case of the serial method. Learning by distributed practice resulted in a 23 percent saving in number of trials required for learning. In the case of paired-associates, however, the advantage is very little (2 percent), the difference in the number of trials required for learning by the two methods lacking statistical significance. Closely similar results are shown in Table 2 where the percentages of subjects for whom distributed prac-

TABLE 1

MEAN NUMBER OF TRIALS REQUIRED TO REACH CRITERION OF ONE PERFECT RECITATION OF SERIAL AND PAIRED-ASSOCIATE LISTS BY MASSED AND DISTRIBUTED PRACTICE. EACH MEAN BASED UPON 64 LEARNING SCORES—2 TRIALS FOR EACH OF 32 SUBJECTS

Serial

	Massed Practice		Distributed Practice
Mean number of trials	12.05		9.23
σ_M	0.48		0.43
Diff.		2.82	
r		+.43	
$\sigma_{diff.}$		0.49	
C.R.$_\sigma$		5.76	

Paired-Associate

Mean number of trials	13.03		12.81
σ_M	0.51		0.58
Diff.		0.22	
r		+.38	
$\sigma_{diff.}$		0.63	
C.R.$_\sigma$		0.35	

TABLE 2

PERCENTAGES OF SUBJECTS FOR WHOM DISTRIBUTED PRACTICE IS SUPERIOR, IS EQUAL, OR INFERIOR TO MASSED WITH SERIAL AND PAIRED-ASSOCIATE LEARNING (32 SUBJECTS)

Serial

	Percent	σ_P
1. Subjects learning more rapidly with distributed practice	62.5	.086
2. Subjects learning equally rapidly with massed and with distributed practice	15.6	.064
3. Subjects learning more rapidly with massed practice	21.9	.073
Difference in percent between 1 and 3	40.6	

Paired-Associate

1. Subjects learning more rapidly with distributed practice	43.8	.088
2. Subjects learning equally rapidly with massed and with distributed practice	18.7	.069
3. Subjects learning more rapidly with massed practice	37.5	.086
Difference in percent between 1 and 3	6.3	

tice is superior, equal, or inferior to massed are given for the two learning methods.

The difference between the paired-associate and serial learning with respect to distribution of practice is brought out clearly in Figs. 1 and 2, where the numbers of trials re-

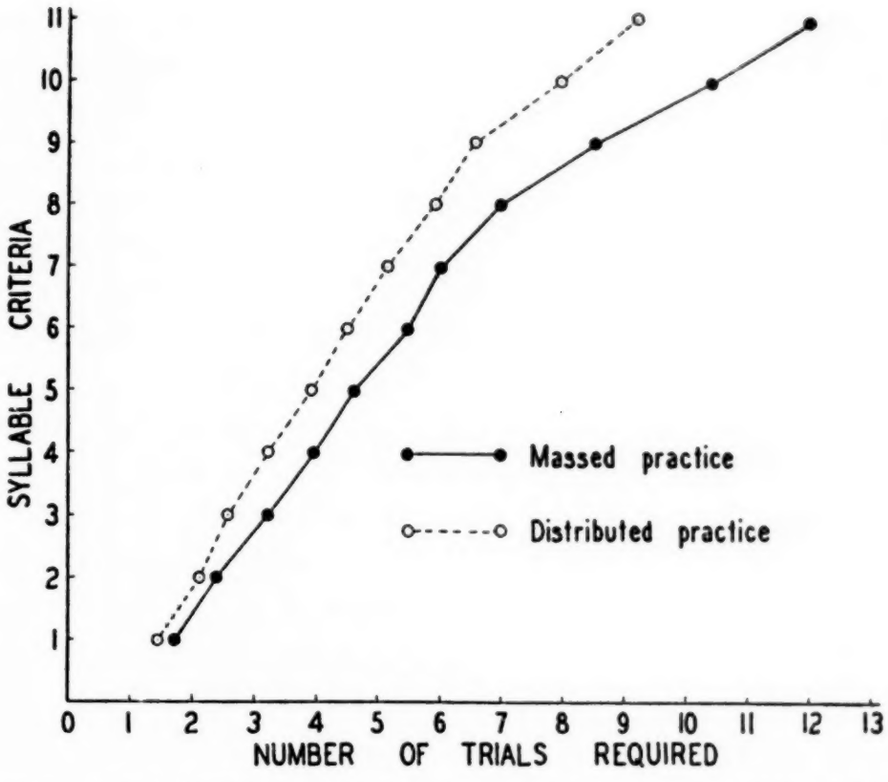


FIG. 1. Graph showing mean number of trials required to reach successive criteria of performance by massed and distributed practice with *serial* learning. (*E.g.* it required an average of 6.6 trials to get 9 syllables correct by massed practice, 8.5 trials by distributed.) Each point is based upon the mean of 64 learning scores (two from each of 32 subjects).

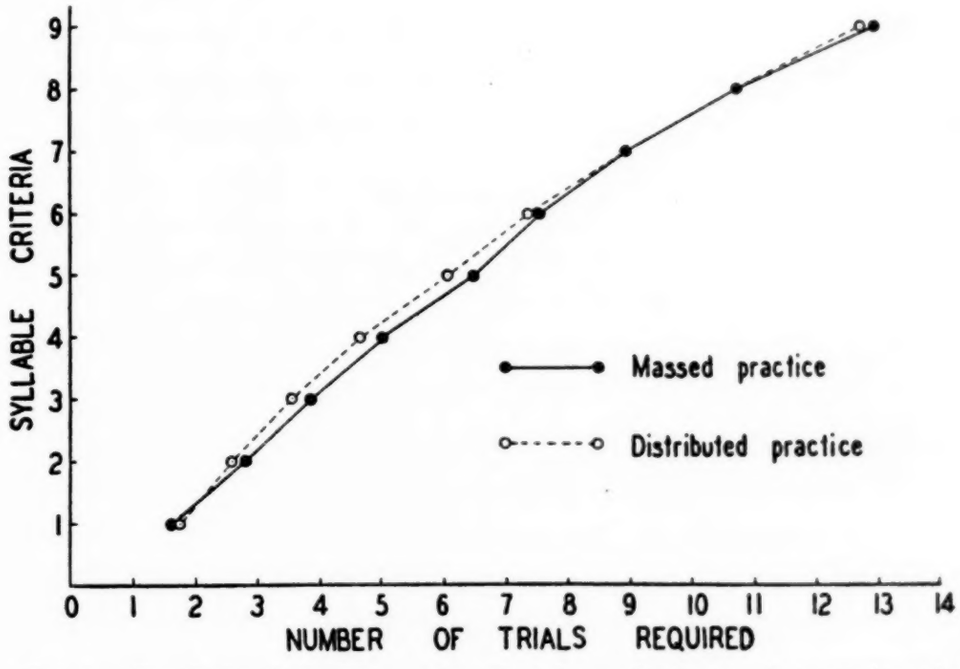


FIG. 2. Graph showing mean number of trials required to reach successive criteria of performance by massed and distributed practice with *paired-assoicate* learning. Each point is based upon the mean of 64 learning scores (two from each of 32 subjects).

quired to reach successive criteria of syllable performance are given. The graph for serial learning is presented in Fig. 1, and for paired-associate learning in Fig. 2. Massed practice is inferior to distributed at every criterion with serial learning, but with the paired-associate material the two curves cross and are practically superimposed. The close agreement in the total number of trials required to reach mastery by massed practice with the two learning methods supports the results of the preliminary equating experiment.

Comparison of the curves obtained with the paired-associate and serial methods shows the former to be more nearly linear than the latter. This is in support of Kjerstad's (**9**) results. He found that additional syllables are progressively more difficult to learn at later stages of learning in the case of serial learning, but not to such a great extent in the case of paired-associates. Another way of stating the decreased rate of learning during the latter phases of learning with serial material is to point out that in the present experiment only about a third (36 percent) of the material was learned during the last half of the massed practice serial-learning, whereas nearly half (46 percent) of the paired-associate materials was learned during the last half of the trials. Distribution of practice with the serial method appears to make the learning curve much more linear, but distribution apparently does not greatly affect the paired-associate curves.

Discussion

Paired-associates have seldom been used in studying the effect of distribution of practice upon the rate of learning. In Jost's (**8**) widely quoted study a paired-associate method of testing was employed, but the experimental variable was not the time between repetitions, as in the present experiment, but the number of repetitions in a block of trials. A constant number of repetitions were given but distributed over a varying number of days. The odd numbered syllables were the stimulus words, and the even numbered syllables the associates to be recalled. He presents his results in terms of the number of correctly recalled associates and their reaction times

Distributed practice increased the number of associates which could be recalled. Since the data are not given in terms of the number of trials required to learn, the results cannot be compared with the present experiment. Similarly, the present results cannot be compared with Lorge's (**11**) experiment employing paired-associates for studying distribution of practice. The material used for memorizing consisted of twelve pairs of three-place numbers as stimuli with two-place numbers as associates. In contrast to the present experiment, a constant time (one minute) was allowed for study of the page upon which the pairs were printed. Another minute was allowed for written recall. A group which was given one minute between trials was only 'slightly and irregularly superior' (p. 26) to an equated group learning by massed practice. Comparison with his results is not possible because of the marked difference in the method of presenting the words and testing recall and because each trial involved the learning of a different list of words.

Lyon (**12**) found the same clear-cut difference between the effect of distributed practice upon learning serial nonsense syllables and upon memorization of poetry and prose that was obtained in the present experiment between serial and paired-associate learning. He attributes the difference to the nature of the connections: "For digits and nonsense syllables, *i.e. material in which there is but little logical connection*,[1] there is considerable saving (by distributed practice)" (p. 91). While it is true that there is a difference in the nature of the logical connection, no explanation is given as to why this difference should be systematically relevant.

Theories advanced for explaining results upon distribution of practice have changed little since the competent review by Warden (**18**). The implication of a number of these formulations has been discussed in relation to paired-associate learning by the writer in analyzing results on reminiscence (**7**). *Rehearsal* is often used to explain the efficacy of distributed practice (**10**). This explanation cannot be applied to the present results since with both the paired-associate and serial

[1] Italics the present writer's.

learning the rest pause during distributed practice was spent in color-naming, an activity which obviates rehearsal.

Differences in the *motivation* (**2**) involved in the learning of the two types of material could similarly only be considered explanatory if such differences had been experimentally demonstrated. There is at present no evidence that there is an increased motivation to learn after a rest pause with serial learning but not after paired-associates. Such an explanation would have a greater probability if, for example, a greater number of trials were required to reach the criterion. But it will be observed that the massed trials were closely equated with respect to this factor. It will also be recalled that the differential effect of the rest-pause upon serial and paired-associate learning appears after even a single repetition (Hovland (**7**)). Essentially the same considerations apply to explanations based upon *fatigue* (**3**).

Perseveration (neural 'setting') is frequently used as an explanatory principle for distribution of practice results. It has been espoused most recently by Woodworth (**20**) who uses the term 'consolidation.' The theory may be regarded either as a neurological explanation, as originally intended by Müller and Pilzecker (**15**), or as merely descriptive. In its neurological aspect the theory has been vigorously attacked, notably by Lashley (**10**). No experimental evidence has been obtained to support the hypothetical persistence of neural activity following learning. Furthermore, if the theory were to explain the present results it would have to be assumed that differences exist in the perseverative effect following serial and paired-associate learning. No experimental confirmation of this additional postulate has been presented. If the theory of perseveration is restricted to be purely descriptive, it becomes circular and of no value (**13**). Thus, the appearance of distributed practice efficacy would be used as evidence of perseveration, and perseveration would then be used to explain the advantage of distributed practice.

In its present form the theory of *stimulus-maturation* (**19**, *cf.* also **16**) is difficult to apply to the present results. It predicts that at certain stages of learning distributed

practice is advantageous because the interpolated time permits maturation to take place; at a later stage and with greater complexity the interpolated time is ineffective. With learning of the type employed in the present experiment the theory gives us no clew as to whether the learning is at a stage where it would or would not be subject to maturation. Furthermore, it would have to be assumed that with serial learning maturation takes place; with paired-associates it does not. While this formulation *describes* the present results, there is no independent evidence of this assumption.

Distributed practice efficacy has also been explained on the grounds that the rest pauses permit recovery from *decremental effects* inherent in the learning (**6**). This theory stresses the close relationship between learning and work performance. In expositions of this explanation the concept of *refractory phase* (**7**) is sometimes employed. Applied to the present results these formulations would say that there were factors in the serial learning which would cause decrement which would not be found in paired-associate learning. Two such factors appear possible: (1) The greater uniformity in the mode of presentation of the serial as compared with the paired-associate lists might cause greater decrement in performance. The successive syllables always appear in the same window and in a constant order, whereas with the paired-associates the stimulus word first appears in the left hand window and subsequently the associate appears in the right hand window. This method of exposure might be less monotonous to the subject and hence cause less decrement. This explanation would be closely allied to formulations based on motivation. (2) The other possibility of decremental differences (after the analogy of *refractory phase*) would be based on the fact that even though paired-associates are presented at a two second rate, the response on the part of the subject only has to be made every four seconds with paired-associates, while with serial materials the response must be made at two second intervals. The more rapid rate of making responses would cause decrement in response. A close connection between the present experiment and an

earlier one concerning the effect of rate of exposure upon distribution of practice (**6**) would then be indicated. A theory of this type would also cover the fact that with massed practice the learning of additional units is progressively more difficult, while with distributed practice the units are more nearly equal in difficulty.

The last principal theory explains the advantage of distributed practice as due to the fact that it permits the *conflicting habits* and *interferences* to be forgotten (**14**). It is assumed that the interfering associations are weaker than the ones which are being learned and that they therefore weaken at a more rapid rate. Thus the *effective strength* of response (strength of correct response minus strength of conflicting response) increases for a short interval of time following learning. This is the *reminiscence* phenomenon. Utilization of successive reminiscence effects would then be the basis of distributed practice efficacy. If this theory were to explain the present results it would have to be assumed that serial associations give rise to stronger interference than paired-associates. It is very plausible that the interferences are in the form of the remote associations which occur in serial learning. Here they are the well-known *forward* and *backward associations*. In paired-associate learning, in contrast, the present technique prevents to a large extent the formation of remote associations by constant rotation of the order in which the pairs are presented. This theory would be broad enough to also cover the results of Lyon (**12**) quoted above, since Thorndike (**17**) and others have shown that remote associates are exceedingly rare in learning of the type exemplified by poetry and prose. Rest pauses would then be relatively ineffective with these materials, which is the result obtained by Lyon. His explanation of his results as due to the *logical connection* between the words in the two situations (see above) would thus be subsumed under conflicting associations, since with strong logical connections the possibility of remote associations would be greatly reduced. The differences in the form of the learning curves for these various types

of material reported by Kjerstad (9) might then be closely related to the amounts of interference present.

From the foregoing survey of theories [2] it appears that a number of the commonest theories such as rehearsal, motivation, and fatigue cannot be applied to the present results. The theories of perseveration and stimulus-maturation are phrased in forms that make them extremely difficult to prove or disprove, although they are descriptively adequate. Decremental factors, principally of the refractory phase type, are doubtless involved in many experiments on distribution of practice and may be partially involved here. But it would appear that the most important factor of difference in the two types of learning is the nature of the interferences or conflicting associations. Further study of this theory appears to be indicated.

Summary

Thirty-two subjects learned both serial and paired-associate lists on successive days by massed and by distributed practice. The anticipation method was used for both types of learning. The number of trials required for learning both types of material by massed practice was equated in a preliminary experiment. During distributed practice a two minute rest pause (during which the subjects named colors) was interpolated between successive learning trials.

Learning by distributed practice resulted in a pronounced reduction in the number of trials required for mastery in the case of serial materials. No superiority of distributed over massed practice was found in the case of paired-associates.

The learning curve obtained with massed practice of serial materials shows that progressively more trials are required to reach successive criteria of performance, but with distributed practice the learning curve is more nearly linear. The curves for *paired-associates* were nearly identical for massed and distributed practice, and both approached

[2] Explanations have also been advanced on the basis of Jost's law (8) and the Zeigarnik effect (21). These are not independent theories, but are results which must themselves be explained by an adequate and comprehensive theory of learning.

linearity to a much greater degree than the curve for massed serial learning. These findings agree with the results of Kjerstad (**9**).

The various theories of distributed practice efficacy are discussed in relation to the present results. Differences in the nature of the decremental factors are probably involved, but it appears that the most likely explanation of the difference in the effectiveness of distribution of practice with the two materials is the difference in the nature of the interferences involved. Long associational bonds which are prominent in serial learning have little opportunity to become formed in the present paired-associate technique. If these give rise to conflicting associations and weaken more rapidly during a short interval than do the correct associations, which are presumably stronger, distribution of practice would naturally be more effective than massed practice.

(Manuscript received June 1, 1939)

References

1. Bunch, M. E., and Magdsick, W. K., The retention in rats of an incompletely learned maze solution for short intervals of time, *J. comp. Psychol.*, 1933, **16**, 385–409.
2. Dashiell, J. F., *Fundamentals of objective psychology*. New York: Houghton, Mifflin, 1928.
3. Guthrie, E. R., *The psychology of learning*. New York: Harper, 1935.
4. Hovland, C. I., Experimental studies in rote-learning theory. I. Reminiscence following learning by massed and by distributed practice, *J. exp. Psychol.*, 1938, **22**, 201–224.
5. Hovland, C. I., Experimental studies in rote-learning theory. II. Reminiscence with varying speeds of syllable presentation, *J. exp. Psychol.*, 1938, **22**, 338–353.
6. Hovland, C. I., Experimental studies in rote-learning theory. III. Distribution of practice with varying speeds of syllable presentation, *J. exp. Psychol.*, 1938, **23**, 172–190.
7. Hovland, C. I., Experimental studies in rote-learning theory. IV. Comparison of reminiscence in serial and paired-associate learning, *J. exp. Psychol.*, 1939, **24**, 466–484.
8. Jost, A., Die Assoziationsfestigkeit in ihrer Abhängigkeit von der Verteilung der Wiederholungen, *Z. Psychol.*, 1897, **14**, 436–472.
9. Kjerstad, C. L., The form of the learning curves for memory, *Psychol. Monog.*, 1919, **26**, No. 116.
10. Lashley, K. S., A simple maze: with data on the relation of the distribution of practice to the rate of learning, *Psychobiol.*, 1918, **1**, 353–367.

11. LORGE, I., Influence of regularly interpolated time intervals upon subsequent learning, *Teach. Coll. Contr. Educ.*, 1930, No. 438.
12. LYON, D. O., The relation of length of material to time taken for learning, and the optimum distribution of time, *J. educ. Psychol.*, 1914, **5**, 1–9, 85–91, 155–163.
13. MCGEOCH, J. A., Studies in retroactive inhibition. II. Relationships between temporal point of interpolation, length of interval, and amount of retroactive inhibition, *J. gen. Psychol.*, 1933, **9**, 44–57.
14. MCGEOCH, J. A., Chapter 13 *in* Boring, E. G., Langfeld, H. S., and Weld, H. P., *Psychology*. New York: Wiley, 1935.
15. MÜLLER, G. E., AND PILZECKER, A., Experimentelle Beiträge zur Lehre vom Gedachtniss, *Z. Psychol., Ergänz.*, 1900, **1**, 1–300.
16. SNODDY, G. S., *Evidence for two opposed processes in mental growth.* Lancaster (Pa.): Science Press, 1935.
17. THORNDIKE, E. L., *The fundamentals of learning.* New York: Bureau of Publications, Teachers College, Columbia University, 1932.
18. WARDEN, C. J., The distribution of practice in animal learning, *Comp. Psychol. Monogr.*, 1923, **1**, No. 3.
19. WHEELER, R. H., AND PERKINS, F. T., *Principles of mental development.* New York: Crowell, 1932.
20. WOODWORTH, R. S., *Experimental psychology.* New York: Henry Holt, 1938.
21. ZEIGARNIK, B., III. Über das Behalten von erledigten und unerledigten Handlungen (Untersuchungen zur Handlungs- und Affekt-Psychologie, K. Lewin, Edit.), *Psychol. Forsch.*, 1927, **9**, 1–85.

STUDIES ON THE PSYCHOPHYSIOLOGY OF BOREDOM: PART 2. THE EFFECT OF A LOWERED ROOM TEMPERATURE AND AN ADDED INCENTIVE ON BLOOD PRESSURE, REPORT OF BOREDOM, AND OTHER FACTORS [1]

BY JOSEPH E. BARMACK, Ph.D.

College of the City of New York

In two previous studies by the present investigator (**2**, **3**), it was demonstrated that a 10 or 15 mg. dose of the antihypnotic, benzedrine sulfate, was effective in allaying the development of an unfavorable attitude to a repetitive task. In one of these studies (**3**) some evidence was obtained suggesting that 60 mgs. of ephedrine hydrochloride had a similar though weaker and more transient effect. Since, under certain conditions, a lowered environmental temperature serves as an antihypnotic, this factor was included for study among those which may influence the development of a bored attitude toward a repetitive task.

The feeling of monotony has been explained by McDowall and Wells (**5**) as a product of the inadequacy of blood circulation to the work being performed. They predicted that a fall or inadequate rise in blood pressure would accompany a feeling of monotony arising out of a bored attitude to a task.

Another purpose of the present investigation was to find the effect of an incentive on blood pressure and heart rate during repetitive work. A sharp rise in blood pressure or heart rate accompanying a sudden change in attitude of boredom and feeling of monotony may be used to support the theory of McDowall and Wells. Minor vascular changes, on the other hand, would be inconsistent with their analysis.

[1] I am grateful to Dr. A. T. Poffenberger, of Columbia University, for the generous assistance which made this study possible.

Data of the latter type would add to evidence already available (**1**, **3**), suggesting that the physiological locus for the feeling of monotony is not entirely peripheral.

Subjects and Procedure

The subjects and basic procedure for the experiment on the effect of lowered room temperature are described elsewhere (**3**). The additional information on procedure relevant to this problem is as follows:

The experiment was performed during the cold winter months. The room temperature was altered by controlling the steam radiator valve and the aperture of the window opening. The obvious disadvantages of this type of temperature regulation are many. Included among the disadvantages is the lack of humidity regulation, exposure to draughts and a wide range of temperature fluctuation. Although temperature readings were taken and adjustments were made at 15 minute intervals, variations of 3.0 degrees C. from the desired temperature were common.

This method of temperature regulation was employed because it was the only one available to us at the time. The average room temperature for all subjects during the control session was 24.7 degrees C., and for the experimental session, 15.6 degrees C. For the experimental session, the temperature was lowered before the subject entered the room.

The same subjects and procedure (with certain modifications herein described) were employed to find the effect of an added financial incentive on some indices of boredom. Essentially, ten male subjects were thoroughly acclimated to the Tycos recording sphygmomanometer, and to the Poffenberger pursuitmeter. Prior to the incentive experiment, each subject had his blood pressure and heart rate measured at least 100 times, on at least 13 different occasions, and each subject had operated the Poffenberger pursuitmeter for a minimum of eight and one-half hours. For the main session, they rested for one-half hour, during which three measures of blood pressure and heart rate were obtained. Then they filled out a subjective rating sheet containing six 9-point scales of feelings or attitudes of bored-interested, relaxed-strained, irritated-pleased, peppy-fatigued, sleepy-wide awake, and attentive-inattentive. At a signal, they operated the Poffenberger pursuitmeter for four fifteen-minute periods, at the end of each of which they filled out a subjective rating sheet. Without interrupting the work, blood pressure and heart rate measures were obtained twice during each period, at the end of five minutes and at the end of ten minutes of each fifteen-minute work period. The data for each period were averaged. Following the first four 15-minute periods of work, the subjects were told the following:

"The best score on the pursuitmeter that you have ever made is X. You have just made a score of Y. For *each* of the four remaining periods that you equal or better X, you will get a quarter, which I shall place here" (point to place on table). "Any questions?"

After filling out the subjective rating sheet at the end of the subsequent period, the subject was told his score and, if earned, given the quarter. The procedure for the remaining four periods was the same unless a subject failed to equal or better his score during the fifth period. If that were the case, he was told he would get 50 cents if he were successful in the sixth period. If he failed in the fifth and sixth, he was told he would get 75 cents if successful in the seventh, etc. The purpose of this procedure was to keep the interest of unsuccessful subjects high.

Because of the difficulty we experienced in pegging the diastolic pressures accurately on the Tycos recording sphygmomanometer, these data will be largely disregarded in the discussion of the results.

Results

A. The Effect of a Lowered Room Temperature

The result of a lowered room temperature on the factors studied appears in Fig. 1. Although small and in some

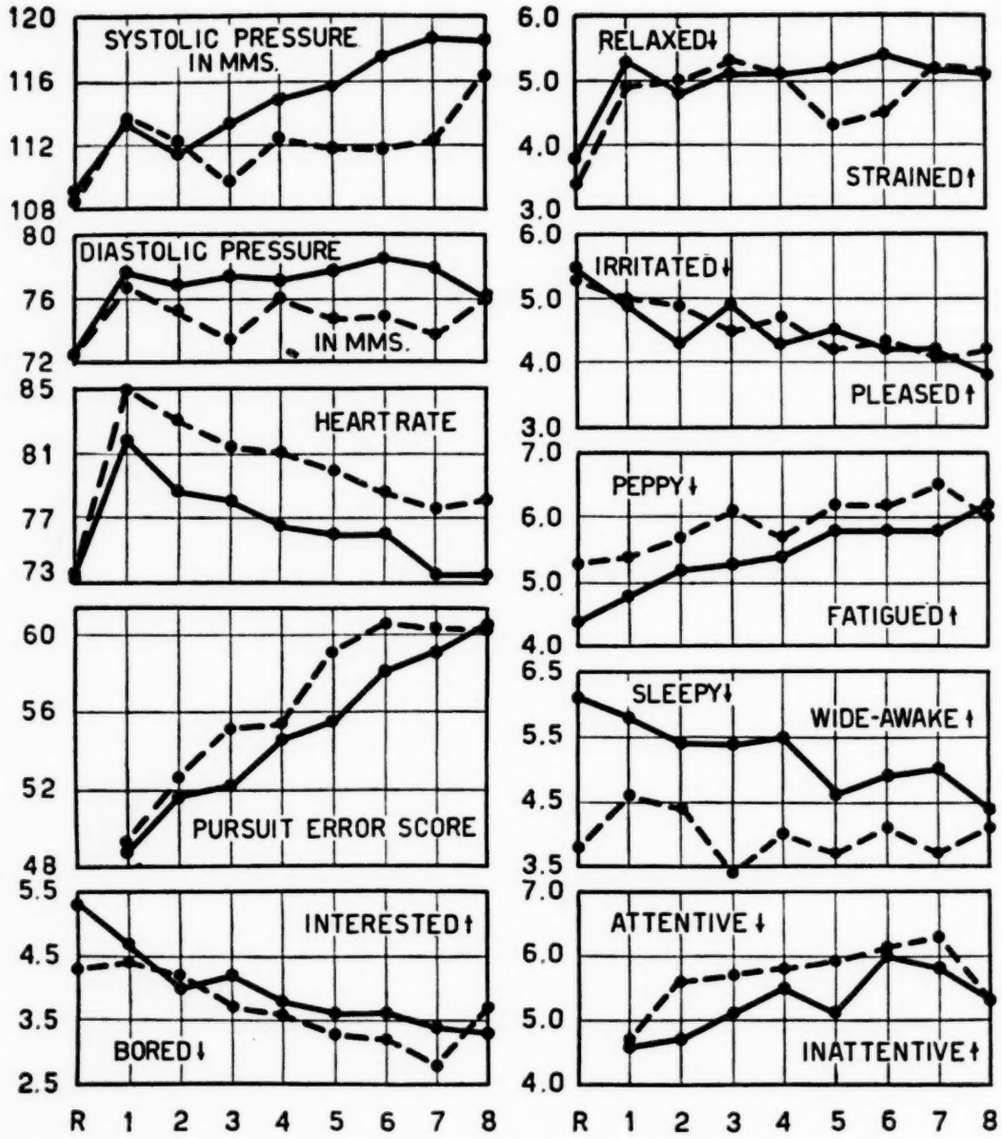


Fig. 1. The effect of a lowered room temperature on ten subjects. Solid lines refer to factors studied in a room temperature averaging 15.6 degrees C. Dotted lines refer to the control condition averaging 24.7 degrees C. *R* represents a thirty-minute rest period. 1, 2, 3, etc., represent successive 15-minute pursuitmeter activity periods. Arrows indicate the direction of increase on the subjective rating sheets.

instances consistent differences are available, most of these differences are to be regarded as barely suggestive, because of their low reliabilities. The systolic pressure, heart rate, and rating on the sleepy-wide awake scale data contain differences which are somewhat more reliable. The p values [2] for mean differences in systolic pressure between the experimental and control sessions at the fifth, sixth, and seventh periods are 8, 5, and 5, respectively. The p values for mean differences in heart rate between the two conditions at the third, seventh, and eighth periods, are 19.0, 7.0, and 2.0, respectively. The p values for mean differences in alertness (sleepy-wide awake) ratings between the two conditions at the end of the rest period, second, fourth, and sixth periods, are 1.0, 19.0, 8.0, and 19.0, respectively.

While most of the p values are not acceptably low, the consistency of the differences should be considered in evaluating their reliability. On this basis, we may conclude that under the conditions of the experiment, the lowered room temperature (15.6 degrees C. as compared to 24.7 degrees C.) resulted in a small increase in blood pressure, a small decrease in heart rate, and an initial, but diminishing, feeling of alertness (sleepy-wide awake rating). The effect of the lowered room temperature on the other factors studied was essentially negative.

B. The Effect of an Incentive on the Factors Studied

The results of the incentive experiment are summarized in Fig. 2. It is apparent that the magnitude of the vascular changes from the non-incentive to the incentive condition is small. On the other hand, the pursuitmeter and most of the subjective rating changes are large. The p values for the mean differences between the data of the fourth and fifth periods are for systolic pressure, 30.0, and for heart rate, 25.

[2] The p value is derived by the use of Fisher's method (4) for determining the reliability of the differences between two means of small samples. The values represent chances in 100 that a difference so large as the one obtained could have been due to chance.

The t value [3] for the mean difference between the fourth and fifth periods in pursuit error score is 6.2; in rating on the bored-interested scale it is 11.3; in rating on the relaxed-strain scale, it is 1.7; in rating on the irritated-pleased scale, it is 3.1;

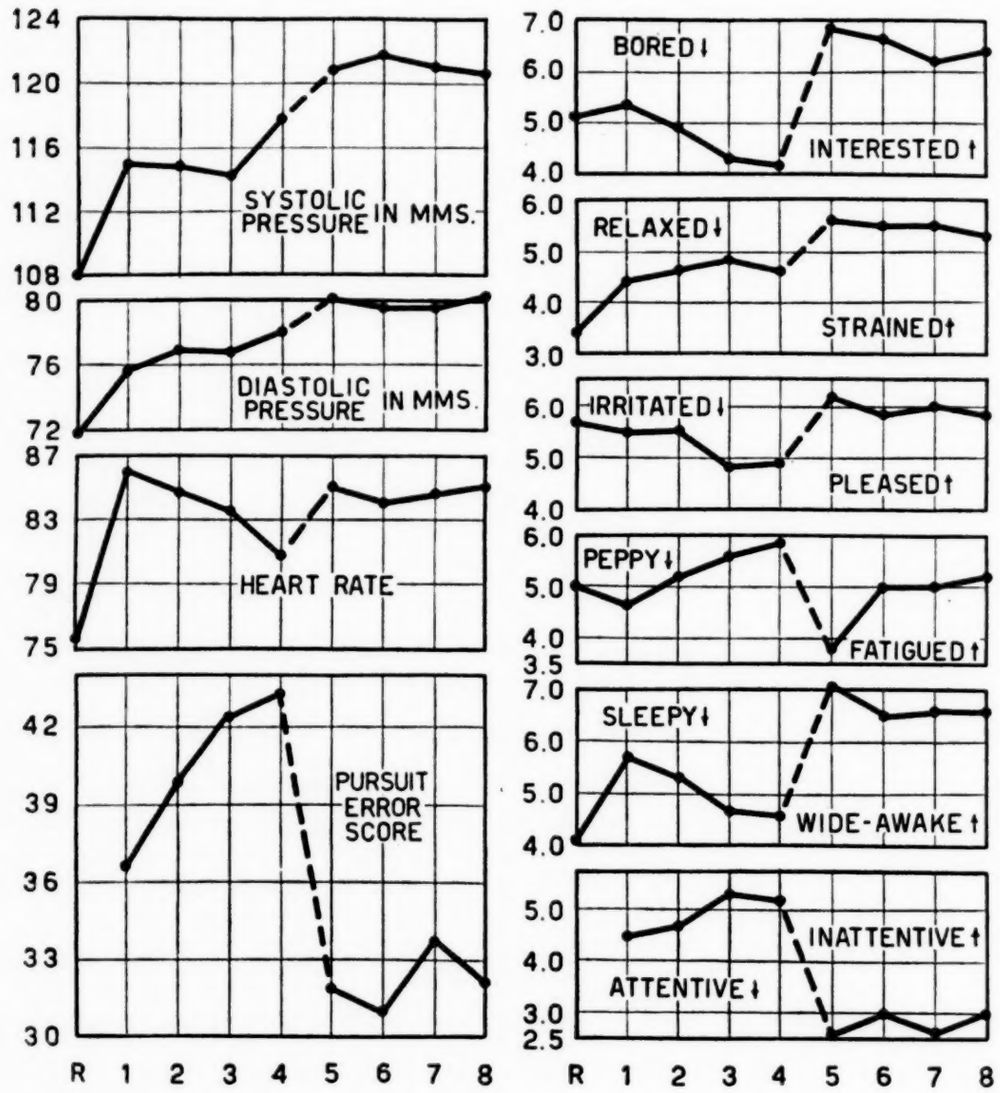


FIG. 2. The effect of an added incentive on pursuit accuracy, report of boredom etc. *R* represents a thirty-minute rest period; 1, 2, 3, etc., represent 15-minute pursuit activity periods. The incentive condition was introduced at the end of the fourth period.

in rating on the peppy-fatigued scale, it is 5.4; in rating on the sleepy-wide awake scale, it is 7.3; and in rating on the attentive-inattentive scale, it is 4.1.

[3] The t value is also derived by Fisher's method. It is analogous to a critical ratio. With ten subjects, when t equals 3.25 or better, the chances are 1, or less than 1, in 100 that a difference as large as the one obtained may be due to chance.

In sum, under the conditions of the experiment, the introduction of an incentive at the middle of two hours of pursuit work resulted in sharp and statistically reliable increases in accuracy and in reports on the subjective scales toward interest, pleased, peppy, wide-awake, and attentive. Small and statistically unreliable increases were obtained in systolic blood pressure, heart rate, and in reports on the subjective rating scale towards 'strained.'

Discussion

A. Temperature Drop

The drop in room temperature from 24.7 degrees C. to 15.6 degrees C. under the conditions of the experiment was practically without influence on the attitude of the subjects toward the repetitive pursuit work. Since this experimental condition resulted in the report of a greater feeling of alertness (rating on sleepy-wide awake scale), than under the control condition, the data here obtained apparently conflict with a theory supported by the writer. According to this theory, boredom is a state of conflict between the tendency to continue and the tendency to escape from a task which has become unpleasant, principally because of inadequate adjustments to it, resulting from inadequate motivation. The adjustments refer to those which vary with the sleep-alertness continuum. Thus, if the lowered room temperature is reported to have an anti-hypnotic effect, the pursuit work should be reported more interesting according to the available indices. The failure of the corresponding change to occur requires discussion.

It is possible that the crude method of temperature regulation, entailing, as it did, uncontrolled humidity and draughts, contributed to an unfavorable work attitude in conflict with the possible stimulating effect of the drop in temperature. Another explanation may rest on the fact that the prolonged exposure to cold may serve as an intrinsic source of discomfort. A third possibility is that the low room temperature is an effective antihypnotic only at the

introduction to it, but loses in this respect as the body becomes adapted to the cold. The consistent decline in report of alertness on the sleepy-wide awake scale supports this interpretation. Under similar experimental conditions, no such decline occurs with the use of 10 or 15 mgs. of Benzedrine. The possibility that the antihypnotic quality of a drop in room temperature may be transient is not inconsistent with the present data. It is not to be expected that such an antihypnotic influence express itself at the beginning of the work period in a particularly higher rate of work or in more favorable *feeling tone reports*. As a result of the relative novelty of the task, and other factors, the subjects may be virtually as well motivated at the beginning of the control as of the experimental condition.

The foregoing hypotheses suggest additional work. A critical experiment would consist of finding the effect of a drop in room temperature of *varying magnitudes* at the *middle* of a two-hour pursuit work session on the factors here reported. The alterations in room temperature should be obtained by appropriate air conditioning equipment.

B. The Incentive Experiment

Under the conditions of the experiment, the introduction of an additional financial incentive for increased pursuit accuracy resulted in only minor blood pressure and heart rate changes, but sharp changes in feeling tone reports indicating a more favorable attitude to the work. The fact that the more favorable ratings were directed primarily toward the work rather than toward the incentive is manifest in the comments of some of the subjects after the experiment. In effect, they were grateful for any incentive that would permit or serve as an excuse for a more dynamic attitude to the task.

The obtained data are damaging to the theory on boredom advanced by McDowall and Wells. According to this theory, the feeling of monotony is a result of the inadequacy of blood circulation in the muscles for the work being performed. They maintain that the physiological manifestation of this

interrelationship should be a drop or inadequate rise in blood pressure. While their theory demands no exact parallelism between blood pressure and feeling tone changes, some degree of concomittance may be inferred.

In view of the added muscular effort associated with the improvement in pursuit accuracy during the incentive condition, the increases of 4 mms. of mercury in systolic pressure and 4 beats per minute in heart rate are negligible. The negligible vascular changes coupled with the striking feeling tone changes add to evidence already available that the physiological locus for the feeling of monotony is not entirely peripheral.

Summary

A. Effect of Lowered Room Temperature

Ten apparatus-acclimated subjects rested for one-half hour during which time three resting measures of blood pressure and heart rate were obtained, using a Tycos recording sphygmomanometer. They then operated a Poffenberger pursuitmeter for eight fifteen-minute periods, at the end of each of which they filled out a subjective rating sheet. Each sheet contained six nine-point scales of bored-interested, relaxed-strained, irritated-pleased, peppy-fatigued, sleepy-wide awake, and attentive-inattentive. Vascular measures during work were taken twice each fifteen-minute period and averaged. The foregoing procedure was followed under two conditions, (*a*) when the room temperature averaged 24.7 degrees C., and (*b*) when the room temperature averaged approximately 15.6 degrees C. Under the conditions of the experiment, the lowered room temperature resulted in the following:

1. A slight increase in systolic pressure.
2. A slight decrease in heart rate.
3. A report of increased alertness (an antihypnotic effect).
4. An insignificant effect on the attitude toward the pursuit work. Further experimental work is suggested by the results.

B. The Effect of a Financial Incentive

The subjects and procedure were essentially the same as in the control session of *A*, but at the end of the first half of the two hours of pursuit work, the subjects were promised a quarter for each of the subsequent four periods in which they equalled or exceeded their best previous score. Scores and rewards were given at the end of each fifteen-minute period. The incentive resulted in:

1. A marked improvement in pursuit accuracy and report of increased interest, alertness, and attentiveness.
2. A relatively small increase in systolic pressure (of the order of 4 mms.) and heart rate (of the order of 4 beats per minute).

The data add to previous evidence that the physiological locus for the feeling of monotony is not entirely peripheral.

(Manuscript received April 27, 1939)

References

1. Barmack, J. E., Boredom and other factors in the physiology of mental effort, *Arch. of Psychol.*, No. 218, 1937, pp. 1–83.
2. Barmack, J. E., The effect of 10 mgs. of benzedrine sulfate (benzyl methyl carbinamine) upon the report of boredom and other factors, *J. Psychol.*, 1938, **5**, 125–133.
3. Barmack, J. E., Studies on the psychophysiology of boredom: Part I. The effect of 15 mgs. of benzedrine sulfate and 60 mgs. of ephedrine hydrochloride on blood pressure, report of boredom and other factors. *This Journal*, 1939, **25**, 494–505.
4. Fisher, R. A., *Statistical Methods for Research Workers*, 1936, p. 125.
5. McDowall, R. J. S., and Wells, H. M., The physiology of monotomy, *Brit. Med. J.*, 1927, **1**, 414.

THE CONVERGENCE OF CORTICAL AND SUBCORTICAL PATTERNS IN MOTOR LEARNING *

BY ROLAND C. TRAVIS

Western Reserve University

INTRODUCTION

In the present study we attempted to isolate experimentally cortical and subcortical functions in the acquisition of a motor skill (eye-hand coördination), and to study the interaction of the lower and higher centers as their functions converge in the learning process.

Simple ocular pursuit of an oscillating target and voluntary manual pursuit of the same target were chosen as representative of a subcortical function and a cortical function respectively.

Evidence that ocular pursuit is primarily subcortically initiated has been obtained from several different lines of investigation. Extirpation experiments by Smith (**4**) in the cat wherein the occipital lobes and striate areas were completely removed leaving the animal blind from the standpoint of general adjustment showed that optic nystagmus was present when a highly illuminated striated drum was revolved in front of the cat. The cat's eyes pursued the moving striations across the visual field and returned with a quick jump. This pattern of ocular behavior is typical of normal optic nystagmus and indicates that without the striate cortex the function of ocular pursuit is still present in subcortical centers.

McGinnis (**3**) found in early infancy that ocular pursuit of a light or bright object was present long before any possibility of conscious recognition or experience could be brought to bear upon this simple act of behavior.

* The author wishes to acknowledge a grant-in-aid from the American Association for the Advancement of Science in partially defraying the expense of this study.

Dodge's experimental differentiation (1) between the various types of eye-movements in the intact adult human indicated that the pursuit movement was a pretty stable neuromuscular pattern resembling a reflex adjustment. Dodge, Travis, and Fox (2) found that ocular pursuit breaks down when the speed of the moving object reaches about sixty to ninety degrees per second. Later Travis (5) found that the quick-phase or saccadic movement reaches a speed of four hundred degrees per second. These two phases of eye-movement seem to be quite distinct experimentally. Furthermore, the pursuit movement cannot be duplicated voluntarily and occurs only when the subject fixates a moving object, or as a result of vestibular stimulation.

The voluntarily initiated function in our experiment was simple manual pursuit of a small target. This function gradually becomes more efficient as learning takes place and it is well known that these coördinated functions of the hand, arm, and shoulder are controlled by the motor area in the cortex.

Experimental Procedure

Figure 1 shows a typical photographic record of (*A*) oscillation of the target, (*B*) manual pursuit of the target, (*C*) ocular pursuit of the target, (*D*) accuracy of manual pursuit and (*E*) time units in 1/33 of a second. The target oscillated through 13 degrees of the visual field at the constant rate of one complete oscillation in approximately 1.5 seconds, and was driven by a synchronous motor through a reducing gear. The target oscillated horizontally across one end of a light tube (55 cm in length) and the subject was seated at the other end in such a way as to permit vision of the target with only the right eye. The left eye served as the recording eye for the Dodge mirror-recorder, which was described in detail in a previous report (6).

The subject's task was to hold the flexible stylus (gravity type) on the oscillating silver target (13 mm in diameter) which was mounted in a panel of bakelite. The details and learning characteristics of this motor function have been given in previous reports (7, 8). Independent lever extensions from the target-panel and the stylus cast shadows across the camera slit thus producing the white oblique lines in Fig. 1. The reflected beam of light from the concave mirror of the Dodge mirror-recorder, which was moved by the eye in pursuing the target, produced the black irregular line in the record. When the stylus was in contact with the target a marker was electrically activated whose shadow was recorded at (*D*) in the record. The time line was produced by a toothed wheel which was rotated across the edge of the camera slit by the camera motor.

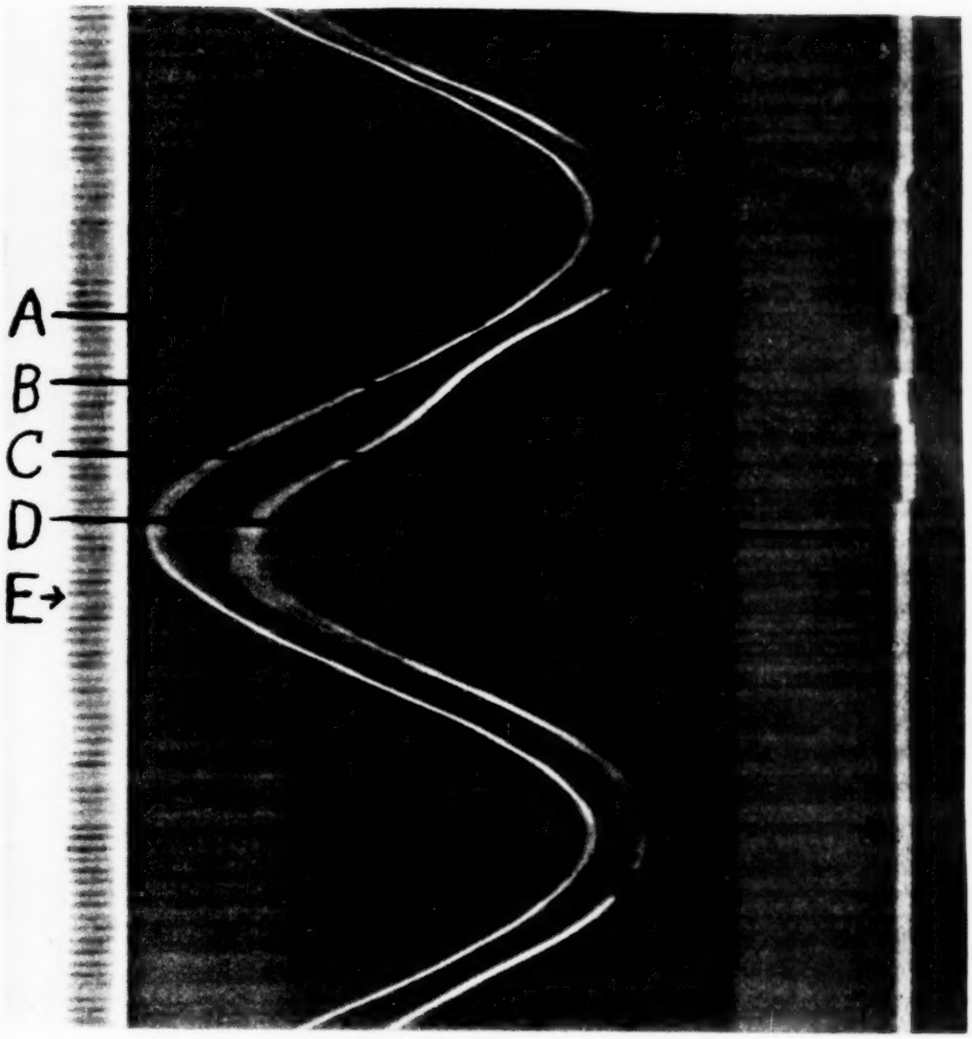


FIG. 1. A typical photographic record of (*A*) oscillation of the target, (*B*) voluntary manual pursuit of the target, (*C*) ocular pursuit of the target, (*D*) coördination accuracy, (*E*) time line in 1/33 of a second.

EXPERIMENTAL RESULTS

Figure 2 shows the relationship between ocular latency, manual latency and accuracy of eye-hand coördination from trial to trial as learning takes place. Latency is given in milliseconds on the left ordinate and coördination-accuracy in percentage of perfect score on the right ordinate. Latency seemed to be the best single measure of ocular and manual pursuit. It represented the time in milliseconds between the onset of movement of the target and the onset of movement of the hand and the eye at the beginning of each oscillation stroke of the target during a continuous series of oscillations.

This graphic representation of the data clearly shows that as the ocular and manual functions become more and more harmonious the accuracy of coördination increases. These data represent averages for four subjects. In the case of one subject the ocular latency and the manual latency were practically identical on the 5th, 6th, and 7th trials and the accuracy of coördination was 90 percent, indicating individual differences in rate of learning and also the possibility of making a

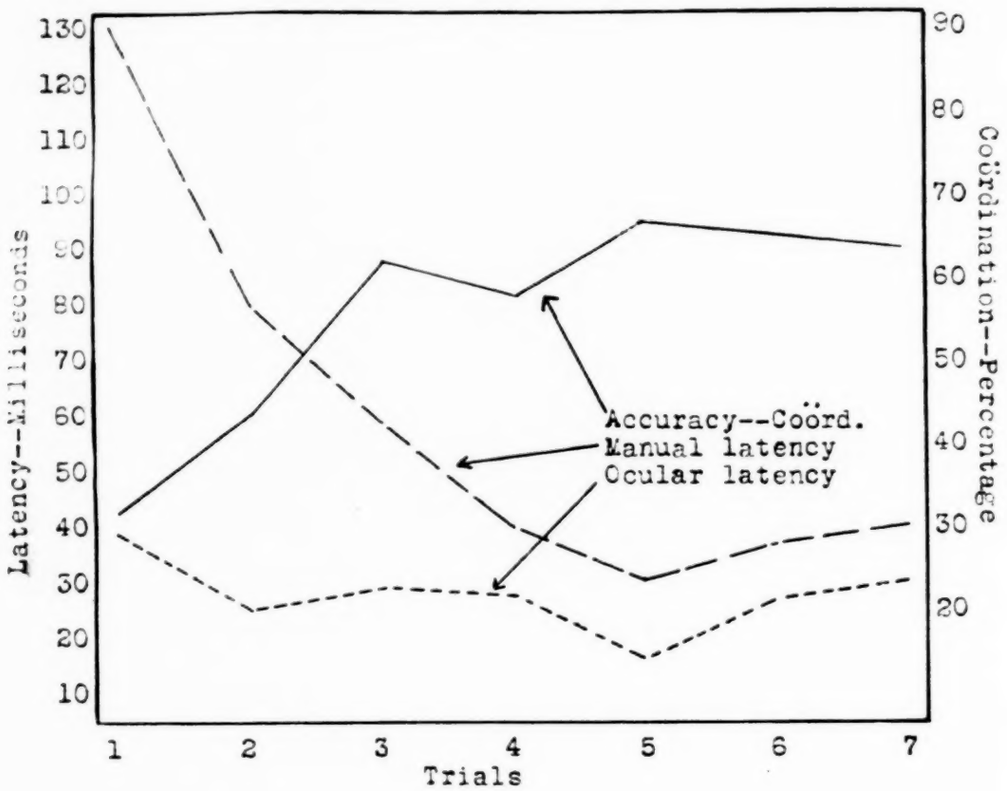


FIG. 2. Relationship between ocular latency, manual latency, and accuracy showing the convergence of cortical and subcortical patterns of behavior in learning in eye-hand coördination (Four subjects).

perfect score as the ocular and manual latencies approach equality.

An outstanding difference between ocular pursuit and manual pursuit in anticipating the target was noted. The line of regard during pure ocular pursuit is rarely observed to be in advance or ahead of the target except when the voluntary or saccadic movement interrupts the pursuit phase. In the case of manual pursuit, anticipatory reactions are

common and occur on the average about 25 percent of the time. This difference is interpreted to mean that pure ocular pursuit is an involuntary subcortically initiated function, whereas manual pursuit is a voluntary cortically initiated function. Another line of evidence bearing on this distinction is that the ocular pursuit latency changed very little from trial to trial, whereas a great change was noted in the manual pursuit latency.

It would be misleading, of course, to maintain that a definite dichotomy exists between cortical and subcortical patterns. No doubt cortical and subcortical patterns overlap in most behavior but there seems to be a fairly definite quantitative difference in terms of whether an act is predominantly cortical or subcortical in character.

A sharp differentiation in speed between the pursuit or slow-phase and the saccadic or quick-phase of eye-movement has been previously reported (**5**) as great as three hundred degrees per second. In this study differences in pausation time between the pursuit-phase and the saccadic or quick-phase were determined. Table 1 shows an outstanding

TABLE 1

SPEED OF OCULAR REVERSAL

Comparison of pausation time of the eye in pursuing the target to a stop and reversal of direction with pausation time of the eye in stopping and reversing direction in voluntary refixation

	Pausation Time of Eye Pursuit Reversal	Pausation Time of Eye Voluntary Refixation
Number of observations	47	57
Mean	35 ms.	254 ms.
S. D.	30 ms.	73 ms.
Range	10 to 60 ms.	120 to 420 ms.

difference between the pursuit-phase and the quick-phase in favor of the pursuit-phase. These data represent further evidence that the quick-phase is a higher level response than the pursuit-phase.

A series of approximately one hundred responses from one subject indicated that the minimum speed of ocular pursuit on the average was about two degrees per second. When

the target moves at speeds lower than two degrees per second ocular pursuit breaks down and an occasional quick-phase movement results.

A great number of measurements (200, 3 subjects) of the latency of ocular pursuit of the target when oscillated by mechanical means were compared with the latency of ocular pursuit when the target was oscillated manually by the subject. In case of the passive ocular pursuit of the target the latency was less and more uniform than for ocular pursuit of the target which was oscillated manually by the subject. The critical ratio (sigma of difference) was 1.2 between the means. This is not a very great difference but indicates a confusion and some disharmony when cortical and subcortical patterns combine in a relatively new task.

Summary of Conclusions

In studying the convergence of a subcortical pattern (ocular pursuit) and a cortical pattern (manual pursuit) as learning takes place from trial to trial the following conclusions seem important.

1. As the latency of ocular pursuit and manual pursuit progressively approached equality, accuracy of coördination progressively increased.

2. Further evidence bearing on the hypothesis that ocular pursuit is primarily subcortical and manual pursuit is primarily cortical in character was found:

a. The latency of ocular pursuit changed very little during the seven trials, whereas the latency of manual pursuit decreased markedly from trial to trial indicating a modifiable function.

b. Anticipatory reactions are frequent in manual pursuit but rare in ocular pursuit.

c. Ocular pursuit of the target oscillated mechanically is more accurate and uniform than ocular pursuit of the target oscillated manually by the subject. This indicates interference in case of combining cortical and subcortical patterns in a new task.

3. The minimum speed of ocular pursuit was found to be about two degrees per second. The maximum speed was found to be sixty to ninety degrees per second (previous study).

4. The pausation time of the eye when the target reverses direction in ocular pursuit was found to be 35 ms. Whereas the pausation time of the eye in voluntary refixation of two points was 254 ms. This outstanding difference probably indicates that ocular pursuit is primarily subcortical whereas voluntary refixation is initiated in the cerebral cortex.

(Manuscript received June 2, 1939)

References

1. Dodge, R., *Elementary conditions of human variability.* New York: Columbia University Press, 1927.
2. Dodge, R., Travis, R. C., and Fox, J. C., Optic nystagmus: III. Characteristics of the slow-phase, *Arch. Neur. and Psychiat.*, 1930, **24**, 21–34.
3. McGinnis, J. M., Eye-movements and optic nystagmus in early infancy, *Genet. Psychol. Monog.*, 1930, **8**, 321–430.
4. Smith, K. U., The postoperative effects of removal of the striate cortex upon certain unlearned visually controlled reactions in the cat, *J. genet. Psychol.*, 1937, **50**, 137–156.
5. Travis, R. C., The latency and velocity of the eye in saccadic movements, *Psych. Monog.*, 1936, **47**, No. 212, 242–249.
6. Travis, R. C., Experimental studies in ocular behavior: I. The Dodge mirror-recorder for photographing eye-movements, *J. gen. Psychol.*, 1932, **7**, 311–327.
7. Travis, R. C., Length of the practice period and efficiency in motor learning, *J. exp. Psychol.*, 1939, **24**, 339–345.
8. Travis, R. C., A manual pursuit-oscillator, *Amer. J. Psychol.*, 1938, **51**, 406–409.

THE VALIDITY OF JUDGMENTS AS A FUNCTION OF THE NUMBER OF JUDGES

BY H. J. EYSENCK

From the Psychological Laboratory, University College, London

In a previous issue of this *Journal*, Dr. K. Gordon has shown that in the discrimination of weights the validity of judgments increases as the number of judges becomes larger (**1**). The question has been raised if a similar relation obtains between judgments of beauty or affective value, and the number of judges (**2**).

The problem, however, is perfectly general; research workers interested in judgments of character, in rankings of essays, or in fact in any investigation involving 'judgments' or rankings of one kind or another, will find it arising in their own fields also.

One great difficulty in experimentation of this kind is that the criterion, or 'true standard,' is not given externally, as in the case of the weights, but has to be deduced from the experimental data themselves.

In order to provide such a criterion in the field of æsthetics, I have carried out a series of experiments on visual æsthetic appreciation, which show that 'a general factor of taste is the most important single determinant for individual æsthetic preferences' (**3**).

It follows that the criterion will simply be the average judgments of a large number of judges: chance errors and 'bipolar factors' will balance out in the long run and leave the correct order, in the same way that in Dr. Gordon's experiment the average judgment of 200 subjects agreed perfectly with the independent criterion, *viz.* the actual weights used.

I was able to obtain 900 rankings, in order of preference, of 12 uncoloured pictures. Of these 900 judgments, 700

were used as the standard, or independent criterion; the remaining 200 constituted the experimental group, corresponding to Dr. Gordon's 200 subjects.

The average orders for the standard group and the experimental group were found to correlate perfectly, making the validity coefficient for the pool of 200 subjects equal to 1.00.

The average of the 200 correlations from the experimental group with the standard order gives us the average validity of one judge at a time. Pooling the judgments of five, ten, twenty, and fifty judges at a time, and correlating the averages of these pools with the criterion, we obtain the validity coefficients for the respective number of judges. The results are shown in the following table (Dr. Gordon's results in brackets):

TABLE I

Number of Rankings Pooled	Average Correlation With Criterion
1	.47 (.41)
5	.77 (.68)
10	.86 (.79)
20	.94 (.86)
50	.98 (.94)

Figure 1 shows the relation between number of judges and

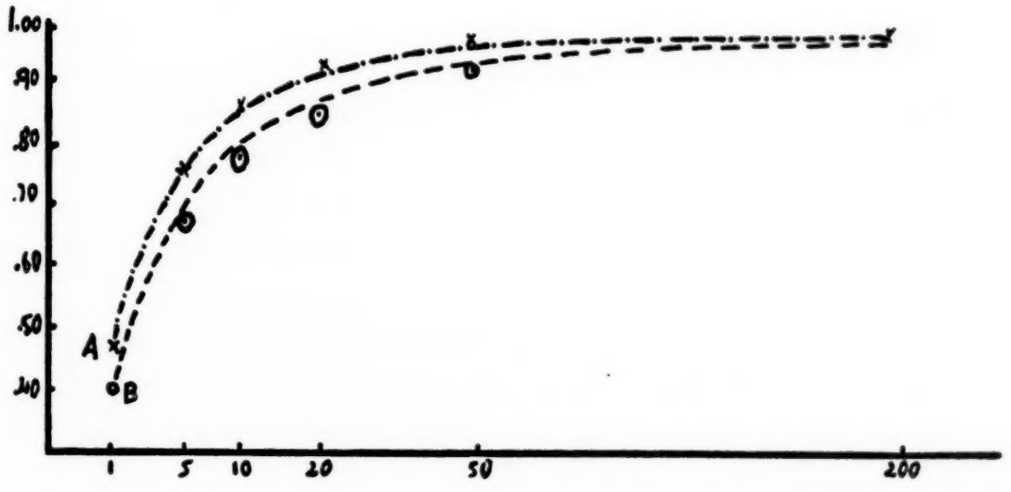


FIG. 1. Relation between number of judges and correlation of their pooled judgments with independent criterion. Curve *A*—æsthetics experiment; curve *B*—weights experiment.

correlation with the criterion for these two investigations. To keep the diagram within reasonable limits, the square root of the number of judges has been taken.

The similarity between the two curves is certainly very marked. It suggests the possibility of finding a formula which would give us the whole curve once the average intercorrelations were known. Such a formula has been derived, in a rather different connection, by Professor C. Burt on purely mathematical grounds (**4**).

His problem was to obtain an estimate of the probable correlation of a set of averaged marks with the true marks, obtained from n examiners, whose average correlation with one another, $\bar{r}_{kk'}$, is known. This formula, being based on reasoning similar to our own, is directly applicable to the present case.

Let $r_{\bar{k}g}$ be the correlation of the average orders with the true order. Let $\bar{r}_{kk'}$ be the average intercorrelation of all the various rankings. Then

$$r_{\bar{k}g} = \sqrt{\frac{n\,\bar{r}_{kk'}}{1 + (n-1)\,\bar{r}_{kk'}}}. \qquad (1)$$

If we want to find the average correlation of one observer at a time with the true order, this reduces to

$$\bar{r}_{kg} = \sqrt{\bar{r}_{kk'}}. \qquad (1a)$$

The extent of agreement of the values given by the formula with those obtained by experiment is shown in Fig. 1. The curves are drawn according to the values given by the formula, and the experimental values are marked by crosses and rings.

Another formula, following directly from the first, enables us to predict how many subjects would be needed to attain a specified degree of accuracy in our results. Using the same notation as before,

$$N = \frac{r_{\bar{k}g}^2(1 - \bar{r}_{kk'})}{\bar{r}_{kk'}(1 - r_{\bar{k}g}^2)}. \qquad (2)$$

Weighting would hardly improve the results which can be very quickly obtained with these formulæ; it has been shown often enough that the unweighted average is practically identical with the weighted average in experiments of this kind.

In practice, it is, of course, unnecessary to compute the intercorrelations ($\bar{r}_{kk'}$) explicitly. Their average can be determined directly by the formula quoted by Kelley (**5**, p. 218, eq. 172).

The following table (Table II) has been compiled to give

TABLE II

TABLE TO DEDUCE THE EXPECTED CORRELATION FROM THE AVERAGE INTERCORRELATION

Average Intercorrelation $\bar{r}_{kk'}$	Correlation With True Order of No. of Rankings					
	1	5	10	20	50	200
.05	.22	.46	.58	.71	.85	.95
.10	.32	.60	.73	.83	.92	.97
.15	.39	.69	.79	.88	.95	.99
.20	.45	.75	.85	.91	.96	.99
.25	.50	.79	.88	.93	.97	.99
.30	.55	.82	.90	.95	.98	.99
.35	.59	.85	.92	.96	.98	1.00
.40	.63	.88	.93	.96	.98	—
.45	.67	.89	.94	.97	.99	—
.50	.71	.91	.95	.98	.99	—
.55	.74	.92	.96	.98	.99	—
.60	.77	.94	.97	.99	1.00	—

the average correlation with the true order of one, five, ten, twenty, fifty, and two hundred judges, and will enable the investigator to judge at a glance which way the results are tending, or how many subjects to use.

It will be noticed that the relation expressed by formula (1a), *i.e.* between the average intercorrelation and the average correlation with the true order, is the same as that between a correlation coefficient and the percentage of causal factors measured ($100\, r_{xy}^2$).

This is so because we may regard the theory underlying the formula for measuring the percentage of causal factors as an

extension of the general factor theory on which this research is based.

(Manuscript received May 31, 1939)

References

1. Gordon, K., Group judgments in the field of lifted weights, *J. exp. Psychol.*, 1924, 7.
2. Guilford, J. P., *Psychometric Methods*. New York, 1936.
3. Eysenck, H. J., General and specific factors in æsthetic judgments. Paper read to the Brit. Psychol. Soc. on 1st April, 1939.
4. Burt, C., *The Marks of Examiners*. London, 1936.
5. Kelley, T. L., *Statistical Methods*. New York, 1924.

Vol. 25, No. 6 December, 1939

193

Journal of Experimental Psychology

DEC 5

EDITED BY
SAMUEL W. FERNBERGER
UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CHARLES W. BRAY, *Princeton University;* ELMER K. CULLER, *University of Rochester;* CLARENCE H. GRAHAM, *Brown University;* JOY P. GUILFORD, *University of Nebraska;* FRANCIS W. IRWIN, *University of Pennsylvania;* DONALD G. MARQUIS, *Yale University;* ARTHUR W. MELTON, *University of Missouri;* CARROLL C. PRATT, *Rutgers University.*

CONTENTS

Time Order Error in Successive Judgments and in Reflexes. I. Inhibition of the Judgment and the Reflex: H. PEAK 535

Goal Gradient, Anticipation, and Perseveration in Compound Trial-and-Error Learning: C. J. HILL 566

Spontaneous Recovery of the Galvanic Skin Response as a Function of the Recovery Interval: D. G. ELLSON 586

Level of Aspiration in Response to a Prearranged Sequence of Scores: J. W. GARDNER 601

Experimental Studies in Rote-Learning Theory. V. Comparison of Distribution of Practice in Serial and Paired-Associate Learning: C. I. HOVLAND 622

Studies in the Psychophysiology of Boredom. II. The Effect of a Lowered Room Temperature and an Added Incentive on Blood Pressure, Report of Boredom, and Other Factors: J. E. BARMACK 634

The Convergence of Cortical and Subcortical Patterns in Motor Learning: R. C. TRAVIS 643

The Validity of Judgments as a Function of the Number of Judges: H. J. EYSENCK 650

PUBLISHED MONTHLY, TWO VOLUMES PER YEAR
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND OHIO STATE UNIVERSITY, COLUMBUS, OHIO

Entered as second-class matter, July 1, 1920, at the post-office at Lancaster, Pa.

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN A. MCGEOCH, *Editor*
State University of Iowa

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears monthly (10 issues), the annual volume comprising about 720 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears monthly (since January, 1937), two volumes per year, each volume of six numbers containing about 625 pages.
Subscription: $14.00 ($7.00 per volume; Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 500 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 560 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

COMBINATION RATES

Review and Bulletin: $11.00 (Foreign, $11.50).
Review and J. Exp. (2 vols.): $17.00 (Foreign, $17.75).
Bulletin and J. Exp. (2 vols.): $18.50 (Foreign, $19.25).
Review, Bulletin, and J. Exp. (2 vols.): $23.00 (Foreign, $24.00).

Subscriptions, orders, and business communications should be sent to

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

THE OHIO STATE UNIVERSITY, COLUMBUS, OHIO